Exhibit 40

PLAINTIFFS' RESPONSE TO DEFENDANTS' MOTION TO EXCLUDE GENERAL CAUSATION TESTIMONY OF PLAINTIFFS' EXPERTS

EXHIBIT
4

Topic	Tab
Infinite scroll	1
AutoPlay	2
Likes	3
Streaks	4
FOMO	5
Algorithms/Features	6
Addictive features (Allcot)	7
Beauty Filters	8
Experimental/Quasi	9
Symptoms and Dx	10
Harms	11
Addiction (Chen/Burke)	12

1

08
scroll

A Survey of Addictive Software Design

CHAUNCEY NEYMAN, California Polytechnic State University

The average smartphone owner checks their phone more than 150 times per day. As of 2015, 62% of smartphone users had used their phone to look up information about a health condition, while 57% had used their phone to do online banking. Mobile platforms have become the dominant medium of human-computer interaction. So how have these devices established themselves as our go to connection to the Internet?

The answer lies in addictive design. Software designers have become well versed in creating software that captivates us at a primal level. In this article, we survey addictive software design strategies, their bases in psychology, and their applications in popular software products. We offer a novel taxonomy to better categorize these addictive design strategies. Additionally, we explore a study conducted at the California Polytechnic State University at San Luis Obispo that illustrates the efficacy of one of the addictive design strategies.

ACM Reference format:
Chauncey Neyman. 2017. A Survey of Addi ... Design. 1, 1, Article 1 (June 2017), 12 pages.

Infinite Scroll

1 INTRODUCTION

As Software becomes an integral part ... se for
the attention of users. This competit ... ntion
strategies that apply psychological p ... fer to
these strategies collectively as "addic ...

Before exploring addictive softwa ... cidate
the importance of the subject. Once ... opular
addictive design strategies. After cov ... dations
of those strategies in psychology. Th ... ategies
in popular applications. Finally, we ... ddictive
design strategies.

2 BACKGROUND

In this section, we will briefly define ... n. These
definitions are meant to provide context for the addictive software design ... sections.

2.1 Human-Computer Interaction

Human-computer Interaction (HCI) is a field that arose in the 1980's as a specialty area in computer science embracing cognitive science and human factors engineering. Today, it is a collection of semi-autonomous fields in human-centered informatics. The field grew in prominence with the rise of personal computers, as the demographic of computer users transitioned from technology professionals to everyday people. [4] Today, Human-computer Interaction remains relevant with the ubiquity of mobile devices. For our purposes, Human-Computer Interaction is the academic topic within Computer Science under which addictive software design falls.

2017. XXXX-XXXX/2017/6-ART1 $15.00

A Survey of Addictive Software Design

CHAUNCEY NEYMAN, California Polytechnic State University

The average smartphone owner checks their phone more than 150 times per day. As of 2015, 62% of smartphone users had used their phone to look up information about a health condition, while 57% had used their phone to do online banking. Mobile platforms have become the dominant medium of human-computer interaction. So how have these devices established themselves as our go to connection to the Internet?

The answer lies in addictive design. Software designers have become well versed in creating software that captivates us at a primal level. In this article, we survey addictive software design strategies, their bases in psychology, and their applications in popular software products. We offer a novel taxonomy to better categorize these addictive design strategies. Additionally, we explore a study conducted at the California Polytechnic State University at San Luis Obispo that illustrates the efficacy of one of the addictive design strategies.

ACM Reference format:
Chauncey Neyman. 2017. A Survey of Addictive Software Design. 1, 1, Article 1 (June 2017), 12 pages.

1 INTRODUCTION

As Software becomes an integral part of the human experience, Software designers compete for the attention of users. This competition has prompted the emergence of several user retention strategies that apply psychological principles to software design. In this paper, we will refer to these strategies collectively as "addictive software design."

Before exploring addictive software design, we must define and outline topics that elucidate the importance of the subject. Once we've established this importance, we will explore popular addictive design strategies. After covering each of these strategies, we will examine the foundations of those strategies in psychology. Then we will present successful applications of these strategies in popular applications. Finally, we will offer our novel taxonomy to better categorize addictive design strategies.

2 BACKGROUND

In this section, we will briefly define and outline topics related to addictive software design. These definitions are meant to provide context for the addictive software design strategies in later sections.

2.1 Human-Computer Interaction

Human-computer Interaction (HCI) is a field that arose in the 1980's as a specialty area in computer science embracing cognitive science and human factors engineering. Today, it is a collection of semi-autonomous fields in human-centered informatics. The field grew in prominence with the rise of personal computers, as the demographic of computer users transitioned from technology professionals to everyday people. [4] Today, Human-computer Interaction remains relevant with the ubiquity of mobile devices. For our purposes, Human-Computer Interaction is the academic topic within Computer Science under which addictive software design falls.

2017. XXXX-XXXX/2017/6-ART1 $15.00

2.2 Psychology

Psychology is the scientific study of how people behave, think and feel. As a science, psychology applies the scientific method to study psychological phenomena. [18] Psychology is the field which provides meaningful scientific explanations to addictive software design strategies.

2.3 Mobile Platforms

Mobile platforms are smart phones and tablets that run software, typically connected to the Internet. They are the dominant medium of Human-Computer Interaction today. As of 2016, 67% of digital time is spent on mobile platforms. [21] This time is increasingly focused in a small subset of apps. Smartphone users spent 45% of their app time on their top app and 73% of their app time on their top three apps. Tablet users spent 87% of their app time on their top three apps. The apps most popular with these users are typically published by Facebook, Google, Snapchat, Amazon, and a few other large publishers. [21] As the dominant medium of Human-Computer Interaction, mobile platforms are the most important platform on which to study addictive software design strategies.

2.4 Internet Addiction

Internet addiction is "a compulsive-impulsive spectrum disorder that involves online and/or offline computer usage and consists of at least three subtypes: excessive gaming, sexual preoccupations, and e-mail/text messaging." The symptoms of each variation of this disorder include excessive use, withdrawal when the computer is inaccessible, tolerance and negative repercussions (including lying, poor achievement, social isolation and fatigue. Some countries, such as China and South Korea, consider Internet addiction one of their most serious public health concerns. [1] Internet Addiction illustrates the real consequences of implementing addictive design strategies.

3 ADDICTIVE SOFTWARE DESIGN STRATEGIES

In the following section, we will explore many of the most popular strategies of addictive design outlined by researchers, bestselling authors, and prominent designers. We will attempt to ground each design strategy with a corresponding psychological study that highlights its efficacy. This list is not exhaustive.

3.1 Variable Rewards

A reward is "something given or received in return or recompense for service, merit, hardship, etc." [16] The brain responds positively to rewards. Rewards become variable rewards when they are given randomly and unpredictably. Variable rewards produce more of the neurotransmitter dopamine than regular rewards. [8] Outside of software design, the method of intermittent variable rewards is used most prominently by slot machines. However, mobile applications have begun to take advantage of this effect through the utilization of notifications and other processes. By intensifying the dopamine surges received by their users, software designers are making their products addictive. [2]

3.1.1 Psychological Study: The Skinner Box.

The psychology of rewards has been studied extensively, especially with regards to the neurotransmitter dopamine. The psychological study most closely tied to intermittent variable rewards involves the "Skinner Box," in which pigeons and rats were conditioned to pull a lever when prompted by a light. Researchers found that dopamine levels in these pigeons and rats surged when they were expecting a reward. These effects were multiplied when treats were rewarded at random; adding variability increased the frequency of the pigeons' completing the intended action. [9]

3.2 Social Reciprocity

Social reciprocity is a "mutual exchange" that is social in nature. [16] We are vulnerable to needing to reciprocate others social gestures. [10] This is illustrated by common etiquette like responding to emails or accepting connection requests. Additionally, as inherently social animals, human beings receive chemical satisfaction when they receive social gratification, such as likes. [9] The highly social components of many popular mobile applications contribute to their addictive properties.

3.2.1 *Psychological Study: The Power of Reciprocity.*

An experiment conducted by Andres Diekmann of the Swiss Federal Institute of Technology explores the power of reciprocity. The experiment involved two groups of test subjects, all anonymous to each other. Subjects from the first group were given 10 tokens worth real money with the option to share some proportion of their tokens with a member of the other group. Later, subjects from the second group were given 10 tokens with the option to share their tokens. Members of the second group reciprocated the gift they'd received almost half the time, and only 10% of the second group did not share their tokens at all. This behavior is not rational, and illustrates the social power reciprocity has on human beings. [5]

3.3 Infinite Scrolling

Infinite scrolling is the idea of loading content on a single page instead of spreading it across a series of pages. [11] It creates an interface through which consuming media is enabled by continuing to scroll, instead of flipping to a new page. This strategy is utilized by many mobile applications. Because there is virtually no end to the materials we can consume via infinite scrolling, we are vulnerable to consuming much more than we would normally without realizing it. This results in users spending much more time on applications than intended.

3.3.1 *Psychological Study: The Bottomless Bowl.*

Infinite scrolling taps into a psychological phenomena illustrated by the "Bottomless Bowl" study. In 2005, Cornell professor Brian Wansink demonstrated that you can trick people into eating more soup by giving them a bottomless bowl. When their soup refills, people will consume 73% more without even recognizing greater feelings of satiation. [10] These findings are consistent with the notion that the amount of food on a plate or bowl increases intake because it influences consumption norms. [3] This suggests that the time sucking power of infinite feeds is derived from their power to normalize uninhibited scrolling.

3.4 The Illusion of Choice

Just as infinite feeds have the power to normalize uninhibited scrolling, Software Designers have the power to control user choices through the layout of their applications. While an application like Yelp appears to empower the users with reviews of nearby restaurants, it's really controlling the limited number of venues users are exposed to. [10] This illusion of choice can keep users engaged for longer, as dissatisfaction with each choice results in the user spending more time browsing alternatives within the application. However, limiting choices the user is exposed to keeps them acting on those options more diligently. [12]

3.4.1 *Psychological Study: Decision Making.*

In an experiment undertaken by Stanford's Mark R. Lepper and Columbia's Sheena S. Iyengar, two separate displays of jams were laid out. One display had 24 different types of jams, while the other had only 6 different types. Of the 242 customers who passed by the extensive display, 60% stopped at the booth with the extensive display while only 40% stopped at the booth with the

limited display. However, only 3% of people who stopped at the extensive display purchased jam while 30% of people who stopped at the limited display bought jam. [12] This is just one of many studies that illustrates the power of limiting choices: while people may be attracted by a large variety of options, they are more likely to act when given fewer choices.

3.5 User Investment

Human beings irrationally project more value on objects they're involved in building or creating. Many applications take advantage of this phenomena by giving users power to curate their own social media profiles. This is supported by the "Ikea effect," wherein consumers were shown to be willing to pay more money for furniture they'd contributed in creating than for pre-built furniture. [9]

Additionally, investing time, data or social capital into a platform causes users to spend more time on that platform. This illustrates the importance of first-to-market principles, as once users have accumulated followers on one platform, they are less likely to leave that platform even if a marginally better alternative with the same functionality arises. [9]

3.5.1 *Psychological Study: The IKEA Effect.*

Named after the popular build-it-yourself furniture chain, the "IKEA effect" refers to the increased value consumers place in something they've had a hand in creating. This phenomena is well documented in a Harvard Business School study titled "The IKEA effect: When Labor Leads to Love." When crafting an IKEA storage box, test subjects were willing to spend over 60% more money for the box they'd built than for a similar box built by somebody else. The effect was exacerbated when subjects were bidding on Origami pieces. This psychological principle has been used in many other applications besides furniture, from instant cake mixes (which became much more popular after consumers were instructed to add an egg) to Build-A-Bear stores. [15]

3.6 Gamification

Closely tied to variable rewards, "gamification" is defined in the tech industry as the process of using game mechanics to reward the completion of tasks. [22] Academically, "gamification" has been defined as "a process of enhancing services with (motivational) affordances in order to invoke gameful experiences and further behavioral outcomes." [14] Experts recommend implementing rewards in small, frequent bits so that the user of an app feels a sense of achievement. They also recommend "sharing loops" that integrate rewards with the users social network by allowing the user to share their accomplishments. [22]

3.6.1 *Psychological Study: Gamification.*

A review of 24 gamification studies found that gamification has a positive impact on the effectiveness of the core service of the platform being gamified. In particular, every study focused on education or learning platforms found a positive effect. In this review, the most commonly implemented gamified elements across the many studies were points, leaderboards, and badges. [14]

4 APPLICATIONS

This section will cover the applications of these addictive design strategies, particularly with regards to some of the most popular mobile apps available. For reference, the four most downloaded iOS apps of all time are (in order) Facebook, Facebook Messenger, Youtube, and Instagram. [6] Other popular apps we'll look at include Twitter, Uber (the driver version), and LinkedIn.

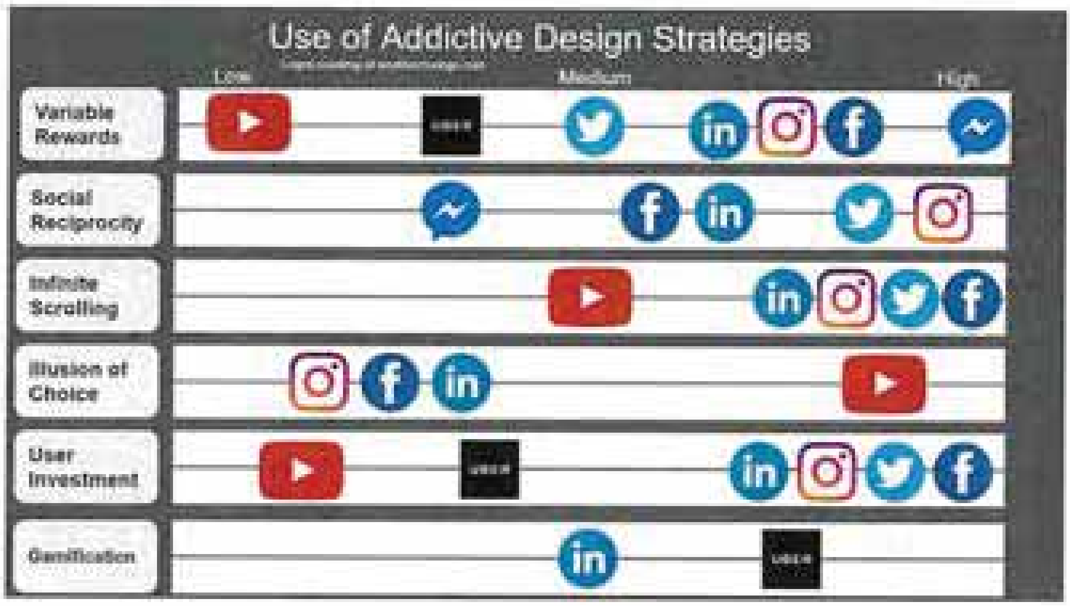


Fig. 1. An illustration of the use of addictive design strategies in popular phone apps

4.1 Variable Rewards

Intermittent variable rewards are used most often in the form of notifications. The Facebook Messenger app relies on notifications and pop ups to alert the user to new messages. The pop ups used by Messenger are some of the most unique of all the major apps, as they appear as bubbles that the user can move around their screen.

The Facebook app utilizes variable rewards in providing notifications for likes, friend requests, and many other activities. Instagram sends notifications for direct messages, when a users' friend posts for the first time in a while, or when somebody likes a users' photo. LinkedIn utilizes notifications similarly to notify users of connection requests, messages and potential job opportunities. Twitter will send users notifications when they are messaged or mentioned. Uber will send drivers notifications when a rider is available to be picked up. And the YouTube app sends notifications when a channel the user subscribes to has posted a new video.

Outside of notifications, Facebook and Instagram are particularly adept platforms at engaging users with "likes." Because a "like" must be given by another user, its delivery is random and satisfying. This is similarly true of "favorites" and "retweets" on Twitter. [9]

4.2 Social Reciprocity

Social reciprocity is either a feature or an emergent property in many social media apps. Apps like Instagram and Twitter have a social etiquette that demands "following back" somebody who has followed you, and the liking and favoriting features can instill a sense of obligation in users to do the same back. Facebook and LinkedIn require friend and connection requests to be accepted before a friendship is made official. And Facebook Messenger messages are shown as "read" to the sending party when opened, motivating users to respond. As these examples illustrate, social reciprocity isn't always an explicit feature. Sometimes it's an unintended consequence of an apps design.

4.3 Infinite Scrolling

Infinite scrolling is most prominently used by Facebook, both on the mobile app as well as the desktop version. It was launched in 2011 at the same time as Facebook Timelines, shortly before Facebook's IPO. [17] The type of infinite scrolling used by Facebook is called "lazy load," because it loads more results as you near the bottom of the page. [13] Before Facebook implemented it, "lazy loading" was used by Twitter and Instagram, and is now used by LinkedIn as well.

YouTube has a different variation on infinite scrolling. While YouTube's search results are paginated (perhaps a result of their ties to Google), they have an autoplay feature which continues to produce related videos to the first one watched manually. Though not explicitly "infinite scrolling," the autoplay feature is a sound example of the same underlying concept of unending content.

4.4 The Illusion of Choice

Most mobile app interfaces use this strategy to direct users between the pages of the apps. What gives these apps the illusion of choice is the way they present a limited set of options as if it were extensive. The YouTube app is particularly adept at this: by presenting the homepage with videos they expect a user to like alongside a search bar, users are given the impression that they can find videos of whatever they want. However, YouTube is meaningfully effecting the content users consume by way of their suggestions and the order of search results.

The illusion of choice is presented effectively in apps like LinkedIn, Facebook and Instagram as well. The results that show up on the main pages of each of these apps appear to be unsorted, chronological posts from all of your friends. However, LinkedIn and Facebook are deliberately tailoring the content that makes it to the top of a users page. And Instagram has recently implemented a similar version of tailored content by showing posts they think you'd like before other posts that are more recent.

Before any of these mobile apps existed, this strategy was being utilized masterfully by Google. While users believe they're searching the full web, Google controls the algorithm that dictates their results. The autocomplete feature of Google (and the similar autocomplete feature used by YouTube) has a powerful effect on the searches users make, yet certain phrases are often blacklisted from those autocomplete lists. For example, typing "crooked hill" on Google will autocomplete with suggestions for restaurants or street names. Typing the same phrase on Bing will autocomplete as "crooked hillary" first - a result that doesn't even appear amongst Google's autocomplete suggestions. [7]

4.5 User Investment

User Investment is often attained through followers, friends and connections. Take Twitter: its core functionality is so easy to recreate that it's been cloned by over 250 other sites. However, Twitter remains dominant in terms of users and valuation. [20] This can be tied to user investment. Twitter users aren't likely to abandon the site for other platforms because they've invested time and energy in gaining followers on their existing account. Even beyond followers, Twitter users have spent time and energy crafting all of the tweets that show up in their history when their profile is examined. [9]

This same notion of user investment is utilized by many other software giants. Instagram uses followers while storing user memories in the form of photos or videos. LinkedIn, as the dominant professional social network, offers access to potential job opportunities through users "connections." Facebook uses friends in addition to their customized Timelines, which store posts as diverse as text, images, and videos. YouTube shows users subscriptions alongside saved videos and videos they've posted. The most successful software apps make it difficult to leave because of the time and energy users invest in them.

Fig. 2. An illustration of our addictive design taxonomy

Facebook has pioneered another form of user investment through their ubiquity on other platforms. Many sites and apps, such as Tinder, Farmville, and GroupMe, allow users to sign up using their Facebook profiles. By making themselves the middle man between users and other apps, Facebook has made it even more difficult for their users to leave.

4.6 Gamification

Uber has historically had issues with retaining drivers. Recently, to counter this retention issue, the Uber app for drivers has been gamified. To keep drivers on the road for a longer period of time each day, Uber sets arbitrary earnings goals that provide achievements when satisfied. They've also started to queue up the next ride for drivers before they've even finished the one they're currently on. [19]

Gamification is exploited more subtly in other platforms. LinkedIn plots the number of people who have viewed your profile alongside goals and suggestions about how to increase your visibility to employers. Health apps, such as MyFitnessPal, set caloric and exercise goals based on your past information. And Snapchat rewards users for "streaks," or days in a row that users have exchanged Snaps. All of these examples gamify processes that are seemingly unrelated to games because gamification taps into the human desire for achievement.

5 CATEGORIZATIONS

In this section, we will create novel categorizations for each the addictive design strategies we've explored. These categorizations are intended to encompass every addictive design strategy presented in this paper in addition to any addictive design strategies implemented by others in the future. These categorizations are rooted in the psychological weaknesses each strategy takes advantage of.

5.1 Craving

Strategies that fall under the "Craving" category take advantage of the physical, chemical response human beings have to desired types of stimuli. Most often the outcome of these strategies manifest in the form of a dopamine rush. For example, intermittent variable rewards such as message alerts and notifications give users a dopamine rush. [10] Similarly, gamified processes such as

achievements can give users a dopamine rush when they are completed. [19] When a user checks their phone expecting a notification, alert, or achievement and they do not receive one, they are illustrating a powerful desire for something, or a "craving." [16]

5.2 Obligation

Strategies that fall under the "Obligation" category take advantage of the human desire for comfort. Human beings naturally seek stability and reassurance from other humans, and addictive design strategies in this category satiate these needs. [10] For example, social reciprocity strategies maintain and uphold existing social norms by validating friendships or inciting correspondence. Similarly, user investment strategies hook users by getting them so used to a platform that leaving that platform would entail leaving friends, family, and a familiar interface. When a user responds to a message or follows back a friend, they are attempting to fulfill an act to which they feel morally bound, or an "obligation." [16]

5.3 Deception

Strategies that fall under the "Deception" category take advantage of human gullibility. This typically entails manipulating a user into doing something they wouldn't normally want to do through the design of an interface. For example, interfaces that utilize infinite scrolling subtly coerce users into spending more time on an app than they intend to. Similarly, giving users an illusion of choice in an app menu while severely constricting their actions to what you want them to do does not always fulfill their desired goals with the app. Each of these strategies gives a mistaken impression to the user, or "deceives" the user. [16]

6 FINAL REMARKS

The well known design strategies outlined in this paper are used by all of the most popular apps on our phones. These software products don't just dominate the market; they dominate our free time. In the future, we anticipate new addictive design strategies to proliferate. Although the specific nature of these strategies may vary, we expect the reason for their effectiveness to remain the same: they take advantage of properties of human psychology. We hope that the categories outlined in this paper increase public understanding of this underlying psychology. Furthermore, we hope this understanding enables fun and responsible software design.

7 APPENDIX

7.1 Experiment: Facebook Scrolling

In this experiment, students were told to download two Google Chrome extensions (add ons to the Google Chrome web browser). One of these extensions ("timeStats") was designed to measure the time they spent on various websites. The other extension ("Stop Scrolling Facebook") was designed to stop the user every 5 minutes to ask if they wanted to continue scrolling on Facebook. The measure of users desire to spend time on Facebook was the difference between the time they spent on Facebook in one week with the "Stop Scrolling Facebook" extension minus the time they spent on Facebook in one week without the "Stop Scrolling Facebook" extension.

7.1.1 Procedure. Students were split into two even groups. The first group was instructed to download both the "timeStats" and "Stop Scrolling Facebook" extensions, while the second group was instructed to only download the "timeStats" extension. After one week, researchers reached out to each group and instructed them on how to report their weekly Facebook activity. The results from each group were recorded (measured to the nearest minute). At the same time, researchers

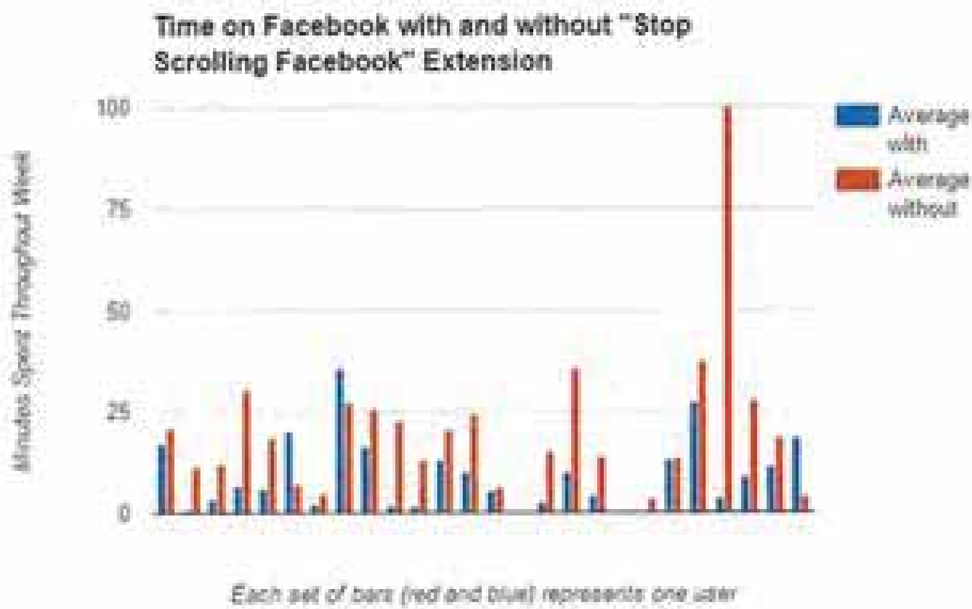


Fig. 3. Results of the Facebook Infinite Scrolling Experiment

instructed the first group to uninstall the "Stop Scrolling Facebook" extension and instructed the second group to install the "Stop Scrolling Facebook" extension. After one more week had passed, researchers reached out again and recorded weekly Facebook activity from each group.

7.1.2 Results. Students spent an average of 10.5 minutes more per week on Facebook without the "Stop Scrolling Facebook" extension than they did with that same extension. The median difference between time spent on Facebook with the extension and without the extension was 9.0 minutes. These differences are significant as students spent an average of only 9.2 minutes on Facebook with the extension vs spending 19.7 minutes on Facebook without the extension - more than twice as long. Further, only 3 students spent more time on Facebook with the "Stop Scrolling Facebook" extension; 21 students spent less time on Facebook with the extension.

7.1.3 Discussion. This experiment was limited by its restriction to tracking Facebook through a browser. When speaking with the students, most admitted that the majority of their Facebook use was conducted through Facebook's mobile apps ("Facebook" and "Facebook Messenger"). However, the difference between time spent on Facebook with and without the "Stop Scrolling Facebook" extension installed is still notable. These results illustrate that users spend more time scrolling the news feed on Facebook than they intend to.

7.1.4 Relevance. This experiment directly tests the efficacy of the "Infinite Scrolling" addictive design strategy outlined in this survey. Its results support the idea that users are deceived into spending more time on a platform with infinite scrolling.

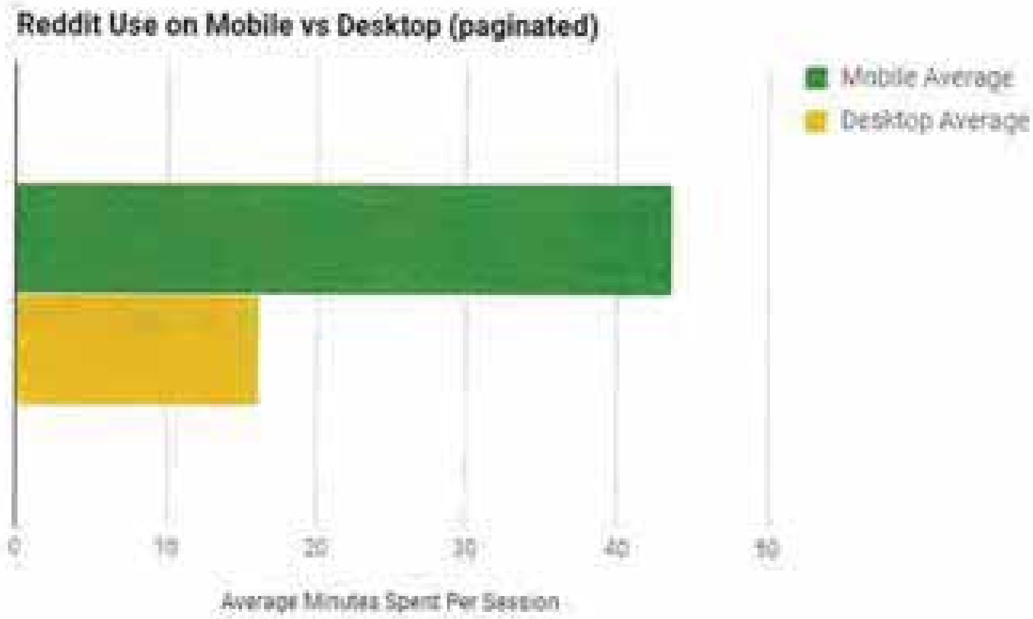


Fig. 4. Results of the Reddit Pagination Experiment

7.2 Experiment: Reddit Pagination

In this experiment, experienced Reddit users were asked to time themselves for two single sessions of Reddit use (where a session is defined as one uninterrupted period of browsing). They were asked to report one session using the Reddit Mobile application (which utilizes bottomless scrolling) and a second session using Reddit via a web browser (which utilizes pagination).

7.2.1 Procedure. Potential test subjects were polled about their Reddit use. First, subjects were asked if they'd used Reddit before. Next, subjects were asked if they had access to the Reddit Mobile application and a browser. After vetting these potential subjects, each subject was asked to self-time and partake in one session of Reddit using the Reddit Mobile application and another session of Reddit using the browser version of Reddit. Subjects were asked to partake in each session on a different day, to avoid seeing the same posts on different days.

7.2.2 Results. Every single test subject spent more time on the Reddit Mobile application than on the browser version of Reddit. Users spent an average of 43.6 minutes on the Reddit Mobile application while they spent an average of 16.1 minutes on the browser version. These averages are very similar to the medians, where users spent a median of 37 minutes on the Reddit Mobile application compared to 15 minutes on the browser version. Qualitative feedback consisted of user comments about which application they felt more or less immersed in, with almost every respondent reporting feeling less immersed on the browser version of Reddit.

7.2.3 Discussion. While this experiment was meant to illustrate the efficacy of bottomless scrolling, these results should be taken with a grain of salt. Because of the nature of the Reddit Mobile application, users almost all used the application on a mobile device while they used the browser version of Reddit on a desktop or laptop device. So while results appear to show that

bottomless scrolling is much more effective, they may just suggest that mobile devices are more compelling mediums for media consumption. Further, the vetting of test subjects (by which we required subjects to be Reddit users with access to both versions of Reddit) severely limited our sample size to a group of 7 people, which is not statistically significant.

7.2.4 Relevance. This experiment is also meant to test the effectiveness of the "Infinite Scrolling" addictive design strategy. As mentioned in the discussion, its results indicate that either infinite scrolling is effective at increasing time spent on an application, users spend more time during sessions on their phone, or some combination of both.

ACKNOWLEDGMENTS

First and foremost, thanks to Dr Clark Savage Turner for introducing the author to academic papers, to LaTeX, and to the idea of writing a survey as a senior project. Thanks to Jenna Provazek, Hanna Yoo, Siddhant Kahal, Thuy Tien and Pankti Gandhi for their contributions to the "Facebook Scrolling" and "Reddit Pagination" experiments. Thanks to Fred Abler for his inspiring insight into addictive technologies in his "User-Centered Interface Design and Development" class. And finally, thanks to Dr Franz Kurfess for the helpful material taught in his "Human-Computer Interaction Theory & Design" class and the opportunity to explore addictive design concepts in greater depth.

REFERENCES

[1] Jerald J. Block. 2008. Issues for DSM-V: Internet Addiction (Editorial). (2008). This editorial summarizes many of the key points of Internet Addiction.
[2] Bianca Bosker. 2016. The Binge Breaker. (2016). This article by the Atlantic covers Tristan Harris' battle against addictive software.
[3] Jill North Brian Wansink, James E. Painter. 2005. Bottomless Bowls: Why Visual Cues of Portion Size May Influence Intake. (2005). This study published in the PubMed journal explores why humans will consume more when they can't accurately gauge portion sizes.
[4] John M. Carroll. 2017. Human Computer Interaction - Brief Intro. (2017). This site gives a formal definition of human computer interaction and its history.
[5] Andreas Diekmann. 2004. The Power of Reciprocity. (2004). This study explores social reciprocity in the context of game theory, and covers a bit of reciprocity's academic history.
[6] Artyom Dogtiev. 2016. Top 10 Most Popular iOS Apps of All Time. (2016). This article reveals the 10 most downloaded iOS applications of all time.
[7] Robert Epstein. 2016. Google Is the World's Biggest Censor. (2016). This US News article highlights the power Google has over Internet traffic.
[8] Nir Eyal. 2012. Want To Hook Your Users? Drive Them Crazy. (2012). This TechCrunch blog post covers some software design strategies outlined in Eyal's book more concisely.
[9] Nir Eyal. 2014. Hooked: How to Build Habit-Forming Products. (2014). This book by Nir Eyal covers some key addictive design methods.
[10] Tristan Harris. 2016. How Technology Hijacks People's Minds. (2016). This essay by Tristan Harris explores how software exploits human nature.
[11] Robert Hoekman and Jared Spool. 2009. Web Anatomy: Interaction Design Frameworks That Work. (2009). This book defines infinite scrolling in a concise manner.
[12] Sheena S. Iyengar and Mark R. Lepper. 2000. When Choice is Demotivating: Can One Desire Too Much of a Good Thing? (2000). This paper explores the drawbacks of excessive choices.
[13] Nathan Johnson. 2012. How does Facebook achieve infinite scrolling? (2012). This Stack Overflow answer describes some of the technical aspects of infinite scrolling as well as where it came from.
[14] Harri Sarsa Juho Hamari, Jonna Koivisto. 2014. Does Gamification Work? A Literature Review of Empirical Studies on Gamification. (2014). This paper reviews empirical studies on gamification.
[15] Dan Ariely Michael I. Norton, Daniel Mochon. 2011. The IKEA Effect: When Labor Leads to Love. (2011). This Harvard Business paper explores the Ikea effect and its business implications.
[16] N/A. 2017. Dictionary.com. (2017). This site is the source of many widely accepted definitions found throughout this paper.
[17] J. Odell. 2011. Once Facebook launches timeline, you'll never want to leave. (2011). This news article covers some of the features of Facebook Timelines shortly after it was released.
[18] University of Queensland. 2017. What is Psychology? (2017). This web page defines psychology and explores how it is studied and researched across the world in academic environments.
[19] Noam Scheiber. 2017. How Uber Uses Psychological Tricks to Push Its' Drivers Buttons. (2017). This NY Times article covers the gamification of Uber from the drivers perspective.
[20] Honey Singh. 2017. Top 250 list of twitter clones sites. (2017). This blog post counts 250 websites with similar functionality as Twitter.
[21] Greg Sterling. 2016. Nearly 85 percent of smartphone app time concentrated in top five apps. (2016). This article covers reports that found users spend 85% of their time on their phone using their top five apps.
[22] Rahul Varshneya. 2014. 6 Techniques to Effectively Gamify a Mobile App. (2014). This blog post outlines gamification methods for app developers.

Nudging Users or Redesigning Interfaces? Evaluating Novel Strategies for Digital Wellbeing Through *inControl*

Alberto Monge Roffarello
alberto.monge@polito.it
Politecnico di Torino
Torino, TO, Italy

Luigi De Russis
luigi.derussis@polito.it
Politecnico di Torino

Infinite Scroll

NUDGING (a)

Figure 1: *inControl* is a browser extension for digital self-control of t... *strategy*, the extension progressively darkens the background as long as... from YouTube). Through a *redesign strategy*, the extension redesign... and proposing minimalistic interfaces to promote intentional use (F... YouTube).

ABSTRACT

As web designers may deliberately adopt design patterns to hook users' attention, researchers and practitioners have innovated several tools for supporting users' digital self-control, hoping to help users self-regulate technology use – especially social networks and video streaming platforms – and achieve digital wellbeing. Unfortunately, these tools often restrict usage, e.g., through self-imposed timers and blockers, limiting interaction possibilities. This paper describes the design, development, and evaluation of two alternative strategies for digital self-control targeting the Facebook and YouTube websites. Specifically, we implemented a Chrome extension that a) highlights when the user is scrolling infinitely by progressively darkening the background (*nudging strategy*), and b) redesigns the homepages isolating guilty pleasure recommendations and proposing a minimalistic interface (*redesign strategy*). We compared the two strategies in a three-week field study with 14 participants, finding that both strategies promoted intentional use and allowed participants to decrease time spent and passive scrolling. In particular, participants liked the nudging strategy more as it supported conscious use without changing the overall user exp... traditional digital self-control tools to diverse approaches that may better support digital wellbeing in the long term.

GoodIT '23, September 06–08, 2023, Lisbon, Portugal

ACM ISBN 979-8-4007-0116-0/23/09...$15.00
https://doi.org/10.1145/3582515.3609523

CCS CONCEPTS

• **Human-centered computing → Social media; Empirical studies in HCI; HCI theory, concepts and models.**

KEYWORDS

technology overuse, attention-capture patterns, digital wellbeing, nudging, commitment interfaces

ACM Reference Format:
Alberto Monge Roffarello and Luigi De Russis. 2023. Nudging Users or Redesigning Interfaces? Evaluating Novel Strategies for Digital Wellbeing Through *inControl*. In *ACM International Conference on Information Technology for Social Good (GoodIT '23), September 06–08, 2023, Lisbon, Portugal.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3582515.3609523

1 INTRODUCTION

In recent years, technology has become an integral part of our daily lives, with several digital services – from social media to video streaming platforms – that help individuals and society for various purposes, from connecting with others to entertainment and educational activities. Nevertheless, there is growing concern about the negative impact these services can have on our digital wellbeing, i.e., a novel psychological construct that defines the challenges of having a good relationship with technology in today's infosphere [10]. In particular, several studies found that users are generally not able to resist temptations of media use [16], thus often

Nudging Users or Redesigning Interfaces? Evaluating Novel Strategies for Digital Wellbeing Through *inControl*

Alberto Monge Roffarello
alberto.monge@polito.it
Politecnico di Torino
Torino, TO, Italy

Luigi De Russis
luigi.derussis@polito.it
Politecnico di Torino
Torino, TO, Italy

NUDGING (a) REDESIGN (b)

Figure 1: *inControl* is a browser extension for digital self-control of the YouTube and Facebook websites. Through a *nudging strategy*, the extension progressively darkens the background as long as the user continues to scroll (Figure 1 (a) shows an example from YouTube). Through a *redesign strategy*, the extension redesigns the interface hiding guilty pleasure recommendations and proposing minimalistic interfaces to promote intentional use (Figure 1 (b) shows the redesigned search-first interface of YouTube).

ABSTRACT

As web designers may deliberately adopt design patterns to hook users' attention, researchers and practitioners have innovated several tools for supporting users' digital self-control, hoping to help users self-regulate technology use – especially social networks and video streaming platforms – and achieve digital wellbeing. Unfortunately, these tools often restrict usage, e.g., through self-imposed timers and blockers, limiting interaction possibilities. This paper describes the design, development, and evaluation of two alternative strategies for digital self-control targeting the Facebook and YouTube websites. Specifically, we implemented a Chrome extension that a) highlights when the user is scrolling infinitely by progressively darkening the background (*nudging strategy*), and b) redesigns the homepages isolating guilty pleasure recommendations and proposing a minimalistic interface (*redesign strategy*). We compared the two strategies in a three-week field study with 14 participants, finding that both strategies promoted intentional use and allowed participants to decrease time spent and passive scrolling. In particular, participants liked the nudging strategy more as it supported conscious use without changing the overall user experience. We conclude with design implications for moving from traditional digital self-control tools to diverse approaches that may better support digital wellbeing in the long term.

GoodIT '23, September 06–08, 2023, Lisbon, Portugal

ACM ISBN 979-8-4007-0116-0/23/09...$15.00
https://doi.org/10.1145/3582515.3609523

CCS CONCEPTS

• **Human-centered computing → Social media; Empirical studies in HCI; HCI theory, concepts and models.**

KEYWORDS

technology overuse, attention-capture patterns, digital wellbeing, nudging, commitment interfaces

ACM Reference Format:
Alberto Monge Roffarello and Luigi De Russis. 2023. Nudging Users or Redesigning Interfaces? Evaluating Novel Strategies for Digital Wellbeing Through inControl. In *ACM International Conference on Information Technology for Social Good (GoodIT '23), September 06–08, 2023, Lisbon, Portugal.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3582515.3609523

1 INTRODUCTION

In recent years, technology has become an integral part of our daily lives, with several digital services – from social media to video streaming platforms – that help individuals and society for various purposes, from connecting with others to entertainment and educational activities. Nevertheless, there is growing concern about the negative impact these services can have on our digital wellbeing, i.e., a novel psychological construct that defines the challenges of having a good relationship with technology in today's infosphere [10]. In particular, several studies found that users are generally not able to resist temptations of media use [16], thus often

falling victim to compulsive behaviors like mindlessly scrolling social media newsfeeds [41] or watching more videos or movies than intended [30, 44].

While users often associate digital wellbeing problems with a lack of self-control [27], a growing body of evidence suggests that such problems do not happen by accident but are deliberately pursued by tech companies [30, 34, 37]. Specifically, a new class of "dark patterns," the so-called Attention-Capture Damaging Patterns (ACDPs) [37], is nowadays used to describe malicious patterns like the possibility of scrolling an interface infinitely or the massive usage of "guilty pleasure" recommendations in social media and video streaming platforms. Such patterns – exploited to maximize users' time spent and interactions and increase advertisements revenue – undermine people's ability to spend time according to their value [29], making them experience a later feeling of regret [37]. Nevertheless, despite this evidence, researchers and practitioners have traditionally adopted generic approaches to support people's digital self-control: Digital Self-Control Tools (DSCTs) [32, 36], in particular, mainly focus on blocking the overall interaction with a distractive app or website, e.g., empowering users to define self-imposed usage timers, often leading to high attrition rates [24]. In this paper, we explore the adoption of two alternative and novel strategies that support people's digital self-control by targeting specific attention-capture patterns without blocking the interaction of the user with the digital service:

- a *nudging strategy*, through which we highlight the presence of ACDPs in the interface to increase users' awareness and trigger conscious decisions and more meaningful usage sessions;
- a *redesign strategy*, through which the interface is redesigned to mitigate the disruptive effects of ACDPs and promote intentional use.

While similar strategies are starting to emerge in the digital wellbeing research area [29, 38], it still needs to be determined how they may impact the use of different platforms and, in particular, which would be preferred by users. To answer these questions, we designed and developed *inControl*, a Chrome extension targeting the Facebook and YouTube websites. We selected the two websites to consider platforms with different purposes and usages (a traditional social network and a video-streaming platform), and because both websites have been traditionally associated with digital wellbeing problems and included in digital wellbeing research [29, 30, 33]. Each strategy implemented by *inControl* targets a specific ACDP adopted by the two websites. The extension implements the *nudging* strategy by highlighting when the user is trapped by the Infinite Scroll mechanism [34, 37]: as long as the user scrolls Facebook's newsfeed or the recommended videos on YouTube's homepage, the extension progressively darkens the background (see Figure 1 (a) for an example on YouTube). The *redesign* strategy, instead, targets the Guilty-Pleasure Recommendations provided by the two websites, e.g., the posts that Facebook algorithmically shuffles in its newsfeed or the video recommendations on YouTube: these recommendations are isolated in separate pages, in order to restructure the homepages of the two websites as minimalistic interfaces that promote the main intentional tasks (e.g., searching for a video or posting something, see Figure 1 (b) for an example on YouTube).

We compared the two strategies implemented by *inControl* in a three-week field study with 14 participants, during which each participant experienced the nudging and redesign versions of Facebook and YouTube after a week of control. Results show that both strategies led to an overall reduction in the time spent by users on the target websites in most cases. For example, highlighting infinite scroll (nudging strategy) resulted in an overall time reduction of 34% on YouTube and 57% on Facebook. Similarly, having a minimalistic homepage (redesign strategy) resulted in a time-spent reduction of 64% on Facebook. Interestingly, the same strategy led users to slightly increase the time spent on YouTube (5%). Furthermore, while both strategies made participants scroll less, we observed an increased number of clicks, hopefully indicating a more intentional and active usage of the two websites. Data from an exit survey also revealed that users particularly appreciated the nudging strategy as it allowed for more informed usage sessions without limiting or changing the website's functionality. On the other hand, some participants said that the redesign strategy impacted their habitual content consumption on the two websites, although they acknowledged that the minimalistic interfaces reduced distractions.

Overall, this paper contributes (1) the design and implementation of two alternative strategies to contemporary DSCTs, i.e., nudging and redesign; (2) the evaluation and comparison of the two strategies demonstrating that both solutions may promote intentional use and impact how users interact with different websites in the wild; and (3) implications for moving from traditional digital self-control tools to alternative strategies that may better support digital wellbeing in the long term.

2 RELATED WORK

2.1 Attention-Capture Damaging Patterns and Digital Wellbeing

In today's attention-based economy [15], technology companies use design and system functionality to take advantage of users' psychological vulnerabilities to capture their attention and increase the amount of time they spend on digital platforms [30, 37], particularly on social media and video-streaming platforms [34]. These mechanisms, called Attention-Capture Damaging Patterns (ACDPs) [37], are similar to the traditional concept of "dark patterns [20]" in that they manipulate users into performing actions that are not in their best interests. ACDPs can take many forms. Generally speaking, Monge Roffarello et al. [37] found ACDPs that may *deceive* users, e.g., by disguising a sponsored content as a regular post from a friend, and other ACDPs that may *seduce* users with short term satisfaction, e.g., the Infinite Scroll or the Pull-to-refresh patterns, which keep users engaged in passive consumption. These and other seductive ACDPs, such as the Guilty-Pleasure Recommendations and Autoplay features of social media platforms, can cause users to sacrifice their sense of control and agency over their attention [29, 30], inducing "zone states [7]" during which users consume content almost unconsciously. In parallel, it is nowadays clear that excessive usage of digital services has the potential to negatively impact people's digital wellbeing from several perspectives, from undermining users' sense of agency and control [30, 43] to creating problems for social interactions [28]. An "addiction debate [27]" is also starting to emerge, with several researchers now suggesting treating

excessive use of technology like smartphones and social media as a real addiction [4, 25]. Such concerns are echoed by several media articles [1, 12, 21], and this growing interest in the digital wellbeing topic makes users often experience a feeling of regret for not being able to control technology use [13].

2.2 Digital Self-Control Tools, Nudging Strategies, and Commitment Interfaces

Concerns around technology overuse and addiction have interested different research communities for many years. So far, HCI researchers and practitioners have dedicated their efforts to supporting people's digital wellbeing and self-control by innovating what Lyngs et al. [32] called "Digital Self-Control Tools (DSCTs)." These tools are external mobile apps or browser extensions that assist users in self-regulating other distractive apps or websites, mainly adopting self-monitoring techniques [36]: through dedicated productivity dashboards, end users can monitor their time spent on their devices and define interventions like usage timers and blockers, e.g., to use Instagram no more than 30 minutes per day. Although DSCTs are becoming popular even as commercial applications – e.g., see Forest [39], which has gathered millions of users [32] – their main limitation is that their interventions focus on *blocking* the overall interaction with a given app or website, without targeting the internal attention-capture mechanisms adopted by the same service [30]. Indiscriminately restricting use, however, has the risk of limiting needed interaction possibilities without solving the problems at their very root [29]. It is therefore not surprising that contemporary DSCTs have been found to be ineffective in the long term [33], mainly because they suffer from a high attrition rate [24].

Given the above issues, some alternative strategies to traditional DSCTs are starting to emerge, with the aim of having contextualized solutions that can target digital wellbeing threats without placing too much burden and restrictions on end users. For example, Purohit et al. [38] suggested using the concept of *digital nudges* to make social media use less addictive. Nudges have been defined by Thaler and Sunstein [40] as changes in the architecture of a system that can be used to steer users' behavior without forbidding or restricting interaction possibilities. A nudge that makes an ACDP "more visible," for example, may be used to trigger conscious decisions and more meaningful usage sessions, thus setting the stage for longer-term systemic changes [36]. Another promising alternative to traditional DSCTs is to *redesign* user interfaces to minimize the negative impacts of ACDPs. The Adaptable Commitment Interfaces proposed by Lukoff et al. [29] are an example of tools that modify or recreate an existing interface to prioritize instrumental use and promote users' sense of agency and control. SwitchTube [29], in particular, is an alternative to YouTube that allows users to activate a focus mode in which recommended videos are hidden. At the same time, users are free to use an explore mode to receive recommendations and browse videos with a lower sense of agency. In this paper, we build on these alternative strategies to further investigate their impacts on different platforms and users' preferences towards these solutions.

2.3 YouTube and Facebook Websites in the Digital Wellbeing Research

Being two of the most widely used websites, YouTube and Facebook have been the subject of numerous studies investigating the impact of technology overuse on digital wellbeing [29, 30, 33]. Furthermore, they have been the target of several strategies for supporting self-control. The SwitchTube app by Lukoff et al. [29] and the "no newsfeed" and "goal reminder" strategies explored by Lyngs et al. [35] are just two examples targeting YouTube and Facebook, respectively. In response to these concerns and research efforts, both YouTube and Facebook have implemented features to promote digital wellbeing and reduce technology overuse. For example, YouTube now includes a "Take a break" feature that reminds users to take a break after a certain amount of viewing time, while Facebook includes a "Time on Facebook" feature that allows users to track their usage and set time limits for specific activities. Despite these efforts, Cho et al. [15] demonstrated that users still consider Facebook and YouTube as a source of distraction that may trigger feelings of regret for being overused. Habitual use of YouTube, in particular, has the potential to negatively impact users' preferences and goals [6], as well as users' sense of agency [30]. Similarly, studies found that patterns of Facebook use may undermine academic performances [42] and promote anxiety, social isolation, and distress [26, 33]. For these reasons, we selected Facebook and YouTube as the two target platform to study the impacts of nudging and redesign strategies on users' self-control.

3 EVALUATING NUDGING VS. REDESIGNING STRATEGIES

We explored the adoption of novel strategies to support people's digital self-control by designing a Chrome extension implementing nudges and interface redesigns targeting specific attention-capture patterns. After selecting two ACDPs operating on Facebook and YouTube, the two authors, together with a master's degree student developing his thesis, used a design process to prototype, build, and pilot *inControl* before testing the extension in the field.

3.1 Preparatory Design Work

3.1.1 Selection of Attention-Capture Patterns. Informed by an established methodology used in previous works, e.g., [34], the two authors – taking the role of HCI experts – conducted an exploratory analysis by manually inspecting the Facebook and YouTube website to define which ACDPs are shared by the two platforms and select a subset of them to target. The need for HCI expertise, in particular, is motivated by the fact that regular users are not able to detect dark patterns in most cases [9, 18]. Table 1 summarizes the results of our analysis, using the typology of 11 ACDPs we extracted in our prior work [37] as a reference.

Not surprisingly [34], the Facebook website turned out to be a container for several ACDPs (9 out of 11), while we found fewer ACDPs (5 out of 11) on YouTube. Such a disproportion can be attributed to the different usages of the two platforms: while YouTube is mainly meant for consuming videos, Facebook is a social media platform that can be used for different purposes, as it includes contents of different natures. For example, we found that the two "gaming" ACDPs, i.e., Playing by Appointment and Grinding, are not directly

Table 1: Presence of Attention-Capture Damaging Patterns (ACDPs) on Facebook and YouTube websites.

Pattern	Description	Facebook	YouTube
Guilty-Pleasure Recommendations	Viral suggestions to increase use time.	✓	✓
Neverending Autoplay	A new video is automatically played when the previous one ends.	✓	✓
Casino Pull-to-refresh	Animated page reload after swiping.	✗	✗
Infinite Scroll	New content is automatically loaded at the end of the page.	✓	✓
Disguised Ads and Recommendations	Ads and suggestions disguised as regular content.	✓	✓
Recapture Notifications	Notifications to make users start a new session.	✓	✓
Playing By Appointment	Users are forced to use a platform at a given time.	✓	✗
Grinding	Users are forced to perform additional tasks.	✓	✗
Attentional Roach Motel	Logging out or canceling an account is purposefully made difficult.	✓	✗
Time Fog	The interface hides information about time spent.	✗	✗
Fake Social Notifications	System notifications disguised as messages from real persons.	✓	✗

Figure 2: A prototype of the nudging strategy targeting YouTube: as long as the user scrolls videos on the home page, the background color progressively darkens.

adopted by Facebook, but are present in social media games,e.g., FarmVille, that can be accessed through the social network.

Of the 5 ACDPs shared by Facebook and YouTube, we selected the following:

- Infinite Scroll for the *nudging strategy*; and,
- Guilty-Pleasure Recommendations for the *redesign strategy*.

We made this choice for two main reasons. First, we considered patterns with different peculiarities: Infinite Scroll is a pattern related to a user's physical interaction, while Guilty-Pleasure Recommendations is a pattern that is more in line with the traditional definition of deceptive UI patterns. Finally, as Infinite Scroll requires a physical interaction, detecting it is technically easy, and the pattern is a good candidate for being highlighted by a nudging mechanism. On the contrary, being part of the user interface, patterns like Guilty Pleasure Recommendations are good candidates for implementing a redesign strategy.

3.1.2 Nudging vs. Redesign Strategies. Besides selecting the target ACDP for each strategy, we conducted multiple prototyping sessions to ideate solutions for the nudging and redesign strategies. Table 2 summarizes the results of this design phase, describing the three conditions we implemented in the *inControl* extension and evaluated in the field study.

Besides the control condition, which is transparent for the users, we ideate specific nudging and *redesign* strategies for the Infinite Scroll and Guilty Pleasure Recommendations ACDPs included on Facebook and YouTube. For nudging the Infinite Scroll pattern, we took inspiration from Anchor [2], an extension that is part of a set of projects for digital wellbeing proposed by Google [3]. Figure 2 shows a prototype targeting YouTube produced during a prototyping session: the idea is to progressively darken the background of the page to let users know "how far" they are going with their scrolling so that they can decide more autonomously when to stop.

For the redesign strategy that minimizes the impact of the Guilty-Pleasure Recommendations pattern, we took inspiration from the SwitchTube app proposed by Lukoff et al. [29], and we decided to restructure and simplify the Facebook and YouTube home pages to promote intentional usage and self of agency. Figure 3, for example, shows a prototype targeting Facebook: we envisioned a minimalistic home page that highlights features related to intentional usage

Table 2: The three conditions – control, nudging, and redesign – we implemented and evaluated through the *inControl* extension.

Website	Control	Nudging	Redesign
Facebook	Normal Facebook interface.	The background color of the Facebook newsfeed becomes progressively darker as long as the user scrolls down the page.	The home page prioritizes features related to intentional usage (e.g., adding a post or chatting with a friend), isolating distractive features like Facebook Watch in separate pages.
YouTube	Normal YouTube interface.	The background color of the recommended video on the YouTube home page becomes progressively darker as long as the user scrolls down the page.	The home page becomes a search-first interface, with distractive features like recommendations isolated in separate pages.

Figure 3: A prototype of the redesign strategy targeting Facebook: the new home page prioritizes "intentional" features (e.g., adding a post or chatting with a friend) while isolating distractive features (e.g., posts from friends and recommended videos of Facebook Watch) behind quick links.

(e.g., adding a post or chatting with a friend), while isolating distractive features (e.g., posts from friends and recommended videos of Facebook Watch) behind quick links.

3.1.3 *Implementation.* We implemented *inControl* by exploiting the Chrome extension APIs[1] to intercept Facebook and YouTube usage and modify their appearance. Furthermore, we used *Firebase*[2] to log users' data during the field deployment. The extension assigned participants to experimental conditions (Table 2) and used a logger to monitor information about how participants interacted with the two target websites. Before starting the field study, the research team internally piloted the *inControl* extension for two weeks. In this phase, we identified and fixed usability issues, e.g., the background color that darkened too quickly, and problems in logging usage data.

3.2 Methods

3.2.1 *Participants.* We recruited participants by exploiting internal mailing lists and snowball sampling. We used an entry survey to recruit participants that a) self-declared a daily usage of the Facebook and YouTube websites greater than 30 minutes and b) used Google Chrome as a browser. Overall, 14 users were eligible for the study and participated in the in-the-wild experiment of *inControl*. On average, participants (10 males and 4 females) were 25 years old (SD = 4.42). All of them were university students enrolled in B.Sc. and M.S. courses held at our university. After screening the participants, we contacted them to provide instructions to install the extension on their computers. Participants had to sign an informed consent form before participating in the study.

3.2.2 *Procedure and Data Collection.* Figure 4 summarizes the procedure adopted in our field experiment. The test was divided into three distinct 1-week phases:

- **Control**: a week during which *inControl* works in the background by merely collecting YouTube and Facebook usage information without implementing any self-control strategy.
- **Nudging**: a week during which *inControl* implements the nudging strategy on Facebook and YouTube by highlighting the Infinite Scroll pattern.
- **Redesign**: a week during which *inControl* implements the redesign strategy on Facebook and YouTube by restructuring the websites' home pages to minimize the impacts of the Guilty Pleasure Recommendations pattern.

[1] https://developer.chrome.com/docs/extensions/reference/, last visited on May 9, 2023
[2] https://firebase.google.com/, last visited on May 9, 2023

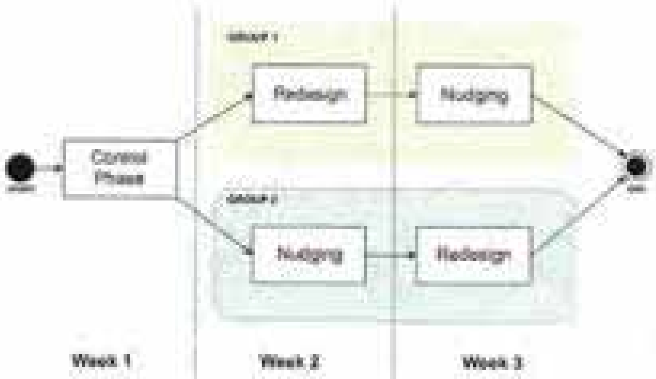


Figure 4: The procedure followed in the field study of *inControl*.

The control phase characterized the first week for all the participants and served as a reference to evaluate the impacts of the two implemented strategies on YouTube and Facebook use. To reduce biases between the tested strategies, participants were randomly split into two groups, with the first one experiencing first the redesign strategy and then the nudging strategy, and the second one experiencing first the nudging strategy and then the redesign strategy. The extension automatically implemented such a process.

3.2.3 Collected Metrics. During the three weeks, we collected different metrics on the impacts of the implemented nudges and redesigned home pages on Facebook and YouTube use. Specifically, *inControl* logged all the users' sessions on the two websites, through which we could calculate metrics like the average daily time spent on a given platform. Furthermore, we also collected the number of clicks performed by users on the two websites and the number of scrolls performed by the users on the interfaces. These two pieces of information were then used to understand whether the presence of a given strategy – besides impacting time spent – also made participants change their interactions with Facebook and YouTube.

At the end of the third week, we also asked participants to fill in an exit survey. In the survey, we asked open-ended questions to investigate users' preferences towards the two proposed strategies, e.g., *"How has the extension changed your use of Facebook and YouTube?"* and *"Which of the two strategies has supported more your sense of agency?"*

3.3 Pre-Registered Hypotheses

In line with best practices adopted in prior work [29], we predefined several specific hypotheses to guide our investigation of how nudging and redesign strategies may influence user experience on Facebook and YouTube. This process helped us define our study protocol in advance and avoid hypothesizing after the results are known [14].

H1: Time Spent. Our first set of hypotheses addressed users' time spent on Facebook and YouTube websites. We expected that the two implemented self-control strategies could reduce the average daily time spent with respect to the control phase, with a more prominent effect of the redesign strategy given its more drastic nature. To summarize, we expected the following on both websites:

- H1a: *time*(*control*) > *time*(*nudging*).
- H1b: *time*(*control*) > *time*(*redesign*).
- H1c: *time*(*nudging*) > *time*(*redesign*).

H2: Interactions. Our second set of hypotheses addressed specific interactions – in terms of the number of clicks and scrolls – with the Facebook and YouTube websites. We expected that the two implemented self-control strategies could reduce the average number of scrolls and clicks per-minute with respect to the control phase. Again, we hypothesized that such an effect could be more evident in the redesign strategy, as the two redesigned home pages contained fewer elements to be scrolled or clicked. To summarize, we expected the following on both websites:

- H2a: *scrolls*(*control*) > *scrolls*(*nudging*).
- H2b: *scrolls*(*control*) > *scrolls*(*redesign*).
- H2c: *scrolls*(*nudging*) > *scrolls*(*redesign*).
- H2d: *clicks*(*control*) > *clicks*(*nudging*).
- H2e: *clicks*(*control*) > *clicks*(*redesign*).
- H2f: *clicks*(*nudging*) > *clicks*(*redesign*).

4 RESULTS

The collected data and the participant's answers to the exit survey allowed us to triangulate the experience of 14 participants in our study using objective and subjective measures.

4.1 Quantitative Results

Table 3 summarizes the usage data collected during the *inControl* field study, reporting the average daily time spent and the average number of scrolls and clicks per minute on Facebook and YouTube in the three study conditions, i.e., control, redesign, and nudging.

As better highlighted in Figure 5, both the nudging and the redesign strategy had a quantifiable impact on the time participants spent on Facebook (a) and YouTube (b) each day on average. Specifically, there was a substantial decrease in the time spent by users on Facebook that can be attributed to the *inControl* extension: compared to the control phase (212.59 minutes spent on average), participants spent 67.42 minutes per day on average (-64%) while using the redesign strategy, while they spent 80.00 minutes on average with the nudging strategy (-57%). These reductions confirm H1a, H1b, and H1c for Facebook. On YouTube, instead, only H1a was verified (287.00 minutes in the control phase vs. 188.80 minutes in the nudging phase, -34%). We instead observed an increased time spent (299.70 minutes, +5%) in the redesign phase that violated H1b and H1c. Although small, a possible explanation for such an increase could be that the YouTube search-first interface allowed users to consume videos more intentionally, thus involving them more than when bombarded with video recommendations on the home page. Similar results have been observed by Lukoff et al. [29] in their evaluation of the SwitchTube app.

For what concerns the influence of *inControl* on users' interactions with Facebook and YouTube, Table 3 shows that the redesigned strategies – resulting in minimalistic home pages with less content to be consumed – led to a significant reduction of scrolls on Facebook and YouTube, thus confirming H2a: 4.72 scrolls

Table 3: A summary of the results of the *inControl* field study, highlighting the average daily time spent and the average number of scrolls and clicks per minute on Facebook (FB) and YouTube (YT) in the control (C), redesign (R), and nudging (N) phases.

	Daily time spent [min]		Scrolls per minute [#]		Clicks per minute [#]	
	FB	YT	FB	YT	FB	YT
C	212.30 (SD = 119.20)	287.00 (SD = 121.90)	274.60 (SD = 138.90)	9.29 (SD = 26.65)	2.84 (SD = 2.61)	0.79 (SD = 1.97)
R	87.42 (SD = 57.20)	299.75 (SD = 113.70)	95.56 (SD = 59.72)	4.72 (SD = 15.54)	4.46 (SD = 2.55)	0.55 (SD = 1.25)
N	88.00 (SD = 57.70)	188.80 (SD = 119.30)	199.80 (SD = 78.06)	7.23 (SD = 27.42)	3.43 (SD = 2.78)	0.70 (SD = 1.77)

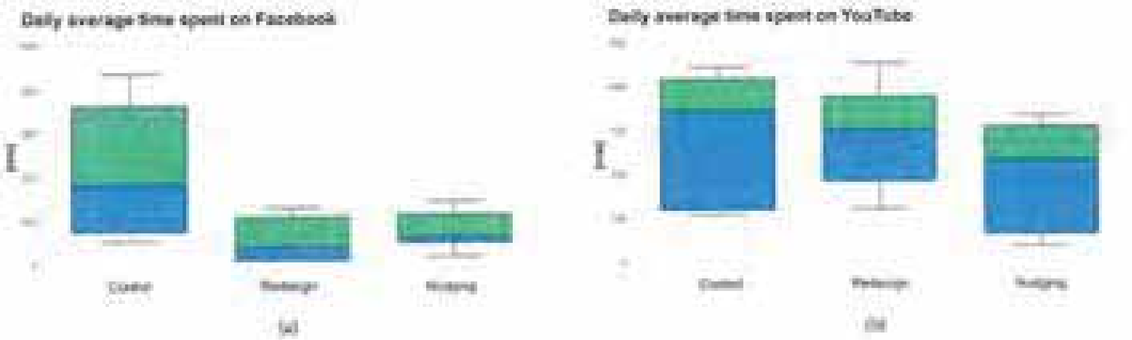


Figure 5: Daily average time spent on Facebook (a) and YouTube (b) in the control, redesign, and nudging phases.

per minute on YouTube (-49% compared to the control phase), and 95.55 scrolls per minute on Facebook (-64% compared to the control phase). On YouTube, we therefore observed an increased time spent and a lower number of scrolls: this finding further points towards a more conscious use of the platform by the participants, as extensive scrolling is typically associated with mindless usage habits associated with a lack of self-control [8]. A smaller reduction in the average number of scrolls on Facebook and YouTube was observed in the nudging phase, too, thus confirming H2b and H2c: 7.21 scrolls per minute on YouTube (-22%) and 199.80 scrolls per minute on Facebook (-27%).

For what concerns the number of clicks, the nudging and redesign strategies resulted in a decrease on YouTube compared to the control phase: 0.55 clicks per minute with the redesign strategy (-31%) and 0.70 clicks per minute with the nudging strategy (-11%). Although these data confirm H2d, H2e, and H2f for YouTube, these reductions, especially those observed in the nudging phase, are smaller than other results reported in this section. The same hypotheses were instead rejected for Facebook. Indeed, there was an increase in clicks both in the redesign phase (4.46 clicks per minute, +57%) and in the nudging phase (3.43 clicks per minute, +21%) compared to the control phase (2.84 clicks per minute). As the time spent on Facebook decreased in the two intervention phases, an increase in the number of clicks suggests a more intentional and active use of the platform.

4.2 Qualitative Results

In the exit survey, participants described *inControl* as useful for having a more conscious use of Facebook and Youtube. Overall, participants said that they gained a greater awareness of their use thanks to the extension, which allowed them to reduce distractions. While participants generally acknowledged that the nudging strategy made them gain awareness *"without removing elements from the screen"* (P4), three users criticized the redesign strategy as it *"reduced the functionality of the websites"* (P3). In particular, a participant said that *"suggestions of new videos is a useful source of content – similar to that I already searched for"* (P11). To mitigate such a problem, two participants suggested the possibility of letting users decide which strategy to activate, in line with the Adaptable Commitment Interface concept proposed by Lukoff et al. [29].

5 DISCUSSION

Our work set the foundations for adopting alternative strategies to support users in self-regulate their usage of digital devices, with a particular focus on Attention-Capture Damaging Patterns (ACDPs) included in social media and video-streaming websites. In this section, we first discuss our findings on two urgent needs in the digital wellbeing research area: promoting awareness rather than blocking use and innovating adaptive self-control solutions. Then, we discuss the limitations of our work and highlight possible future directions.

5.1 Promoting Awareness Rather Than Blocking Use

As multiple studies and reviews have recently demonstrated, one of the main limits of contemporary solutions for supporting people's digital self-control, i.e., DSCTs [32], is that they mainly focus on making users reduce the time they spend on their digital devices [28, 30, 36]. Commercial DSCTs included in the Android [19] and iOS [5] operating systems, as well as research artifacts like NUGU [25] and Lock n' LoL [22], pursue such a goal by letting users define

self-imposed timers that block the interaction with a given app or website after a given amount of time. Applying an intervention that limits the usage of an entire website or app, however, may also interfere with the features of that website or app that are not necessarily a threat to the users' digital wellbeing [29]. At the same time, such a strategy does not address the root cause issues, i.e., the internal ACDPs within the interface [30].

We deemed the findings of our study particularly important, as they contribute to the growing debate that asks digital wellbeing researchers to move beyond traditional DSCTs and screen time metrics [11, 17], considering *"quality of time, not just quantity"* [29]. On the one hand, indeed, the *nudging* and *redesign* strategies implemented by the *inControl* Chrome extension allowed participants to reduce the time they spent on Facebook and YouTube in most cases and reduced scrolling behaviors. At the same time, the redesign strategy made participants slightly increase the time spent on YouTube (5%) and resulted in a higher number of clicks on Facebook (+57%) – even if the time spent on the social network was lower on average compared to the control phase. Overall, this suggests that besides helping users to exercise self-control, having a minimalistic interface allowed users to interact with YouTube and Facebook more intentionally. While Lukoff et al. [29] already demonstrated this tendency on YouTube, our work shows that similar findings also apply to "traditional" social media like Facebook.

5.2 The Need for Adaptive Self-Control Solutions

While both the tested strategies demonstrated the ability to impact usage behaviors, participants expressed their concerns about having solutions that drastically change the appearance of a user interface, demonstrating a preference for the nudging strategy. Nudges that make ACDPs "more visible" could allow users to understand and recognize why using platforms like YouTube and Facebook is so compelling to the point of creating compulsive behavior. As partially demonstrated by our empirical results, such a strategy may also promote meaningful and intentional usage sessions, with the advantages of being less intrusive, i.e., without restricting nor removing interaction possibilities.

That being said, quantitative results of the redesigned Facebook and YouTube interfaces demonstrate that such a strategy is undoubtedly promising. Consequently, our work also echoes findings and discussions about the need for adaptable and adaptive self-control solutions included in a recent work by Lukoff et al. [29].

Adaptable self-control solutions may allow users to activate a given strategy manually, e.g., adding a nudge to the interface or switching to an alternative design, depending on factors like their current intention or mood. While Lukoff et al. [29] explored an adaptable interface for the mobile app of YouTube, the effectiveness of such an approach for different strategies, devices, and target platforms still needs to be determined.

Adaptive self-control solutions may be the next step: instead of asking users to manually switch between strategies – a task that may be challenging – an adaptive DSCT could automatically adopt a strategy that looks promising for a given user in a given context, e.g., adopting personalized prediction models. HCI researchers have already demonstrated the feasibility of analyzing usage data and predicting when a user is trapped in a passive usage session with a device or when the same user has a specific, intentional goal guiding the session [31]. Ideally, user interfaces could provide users with higher-control mechanisms when they have a specific intention and lower-control mechanisms when they have a non-specific intention, e.g., "forcing" a search-only interface for instrumental use [30].

5.3 Limitations and Future Work

Our work has potential limitations. The in-the-wild experiment of *inControl* involved a relatively small group of university students and lasted three weeks, only. These choices and numbers align with most previous experiments about tools for digital self-control: previous work [36] – for example – has found that the average duration of DSCTs experiments is 21 days, with a prevalence of young university students involved. However, we stand with the suggestions provided by the same previous work [36] acknowledging the need to test the generalizability of our findings and proposed self-control strategies in larger and longer studies that involve a varied population.

Furthermore, our experiment followed a within-subject approach through which each participant experienced a control phase and all the intervention strategies. As such, we must acknowledge that some effects of the nudging and redesign strategies may have been circumstantial. Further, between-subject experiments with control groups may be needed to confirm or refute our findings.

Finally, future works could also be conducted to empirically compare effects and users' preferences between the alternative strategies proposed in this work and traditional DSCTs.

6 CONCLUSIONS

This work attempted to take a step beyond restrictive and screen-time-based DCSTs in two main directions: targeting the root causes of technology overuse problems, i.e., the attention-capture patterns exploited by current digital services, and promoting awareness and meaningful use rather than indiscriminately blocking the user's interaction. To this end, we designed, developed, and evaluated *inControl*, a Chrome extension implementing two strategies for digital self-control targeting the Facebook and YouTube websites: a nudging strategy to promote awareness of the Infinite Scroll pattern and a redesign strategy to mitigate the effects of Guilty-Pleasure Recommendations. Results of a field study involving 14 participants showed evidence that both strategies may impact usage behaviors of Facebook and YouTube while promoting intentional and active use, thus demonstrating the possibility of moving from traditional DSCTs to alternative strategies that may better support digital wellbeing in the long term.

ACKNOWLEDGMENTS

The authors want to thanks the all the participants of the study for their availability, and Fabio Stabile who helped with the creation of the *inControl* browser extension as part of his M.S. thesis.

REFERENCES

[1] 2016. Technology is diminishing us. https://www.theguardian.com/books/2016/dec/03/jonathan-safran-foer-technology-diminishing-us Accessed: 2023-05-09.
[2] 2020. Anchor. https://experiments.withgoogle.com/anchor Accessed: 2023-05-09.

[3] 2023. Experiments with Google. https://experiments.withgoogle.com/ Accessed: 2023-05-09.

[4] Cecilie Schou Andreassen, Torbjørn Torsheim, Geir Scott Brunborg, and Ståle Pallesen. 2012. Development of a Facebook Addiction Scale. *Psychological Reports* 110, 2 (April 2012), 501–517. https://doi.org/10.2466/02.09.18.PR0.110.2.501-517

[5] Apple. 2020. Use Screen Time on your iPhone, iPad, or iPod touch. https://support.apple.com/en-us/HT208982 Accessed: 2023-05-10.

[6] Janarthanan Balakrishnan and Mark D Griffiths. 2017. Social media addiction: What is the role of content in YouTube? *Journal of behavioral addictions* 6, 3 (Sep 2017), 364–377. https://doi.org/10.1556/2006.6.2017.058

[7] Amanda Baughan, Mingrui Ray Zhang, Raveena Rao, Kai Lukoff, Anastasia Schaadhardt, Lisa D. Butler, and Alexis Hiniker. 2022. "I Don't Even Remember What I Read": How Design Influences Dissociation on Social Media. In *CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 18, 13 pages. https://doi.org/10.1145/3491102.3501899

[8] Mohammed Bedjaoui, Nadia Elouali, and Sidi Mohamed Benslimane. 2018. User Time Spent Between Persuasiveness and Usability of Social Networking Mobile Applications: A Case Study of Facebook and YouTube. In *Proceedings of the 16th International Conference on Advances in Mobile Computing and Multimedia* (Yogyakarta, Indonesia) *(MoMM2018)*. Association for Computing Machinery, New York, NY, USA, 15–24. https://doi.org/10.1145/3282353.3282362

[9] Kerstin Bongard-Blanchy, Arianna Rossi, Salvador Rivas, Sophie Doublet, Vincent Koenig, and Gabriele Lenzini. 2021. "I Am Definitely Manipulated, Even When I Am Aware of It. It's Ridiculous!" - Dark Patterns from the End-User Perspective. In *Designing Interactive Systems Conference 2021* (Virtual Event, USA) *(DIS '21)*. Association for Computing Machinery, New York, NY, USA, 763–776. https://doi.org/10.1145/3461778.3462086

[10] Christopher Burr, Mariarosaria Taddeo, and Luciano Floridi. 2020. The Ethics of Digital Well-Being: A Thematic Review. *Science and Engineering Ethics* (2020), 2313–2343. https://doi.org/10.1007/s11948-020-00175-8

[11] Marta E. Cecchinato, John Rooksby, Alexis Hiniker, Sean Munson, Kai Lukoff, Luigina Ciolfi, Anja Thieme, and Daniel Harrison. 2019. Designing for Digital Wellbeing: A Research & Practice Agenda. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3290607.3298998

[12] Rory Cellan-Jones. 2018. Confessions of a smartphone addict. https://www.bbc.com/news/technology-44972913 Accessed: 2021-07-15.

[13] Hyunsung Cho, DaEun Choi, Donghwi Kim, Wan Ju Kang, Eun Kyoung Choe, and Sung-Ju Lee. 2021. Reflect, Not Regret: Understanding Regretful Smartphone Use with App Feature-Level Analysis. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 456 (oct 2021), 36 pages. https://doi.org/10.1145/3479600 ✓.

[14] Andy Cockburn, Carl Gutwin, and Alan Dix. 2018. HARK No More: On the Preregistration of CHI Experiments. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3173715

[15] Thomas H. Davenport and John C. Beck. 2001. *Attention Economy: Understanding the New Currency of Business*. Harvard Business School Press.

[16] Liam Delaney and Leonhard K. Lades. 2017. Present Bias and Everyday Self-Control Failures: A Day Reconstruction Study. *Journal of Behavioral Decision Making* 30, 5 (2017), 1157–1167. https://doi.org/10.1002/bdm.2031

[17] Michael A. Devito, Ashley Marie Walker, Jeremy Birnholtz, Kathryn Ringland, Kathryn Macapagal, Ashley Kraus, Sean Munson, Calvin Liang, and Herman Saksono. 2019. Social Technologies for Digital Wellbeing Among Marginalized Communities *(CSCW '19)*. Association for Computing Machinery, New York, NY, USA, 449–454. https://doi.org/10.1145/3311957.3359442

[18] Linda Di Geronimo, Larissa Braz, Enrico Fregnan, Fabio Palomba, and Alberto Bacchelli. 2020. UI Dark Patterns and Where to Find Them: A Study on Mobile Applications and User Perception *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376600

[19] Google. 2020. Our commitment to Digital Wellbeing. https://wellbeing.google/ Accessed: 2023-05-10.

[20] Colin M. Gray, Yubo Kou, Bryan Battles, Joseph Hoggatt, and Austin L. Toombs. 2018. The Dark (Patterns) Side of UX Design. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3173574.3174108

[21] Tristan Harris. 2017. How Technology Hijacks People's Minds. https://www.huffpost.com/entry/how-technology-hijacks-peoples-minds_b_10155754 Accessed: 2021-02-12.

[22] Minsam Ko, Seungwoo Choi, Koji Yatani, and Uichin Lee. 2016. Lock N' LoL: Group-based Limiting Assistance App to Mitigate Smartphone Distractions in Group Activities. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. ACM, New York, NY, USA, 998–1010. https://doi.org/10.1145/2858036.2858568

[23] Minsam Ko, Subin Yang, Joonwon Lee, Christian Heizmann, Jinyoung Jeong, Uichin Lee, Daehee Shin, Koji Yatani, Junehwa Song, and Kyung-Mee Chung. 2015. NUGU: A Group-Based Intervention App for Improving Self-Regulation of Limiting Smartphone Use. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Vancouver, BC, Canada) *(CSCW '15)*. Association for Computing Machinery, New York, NY, USA, 1235–1245. https://doi.org/10.1145/2675133.2675244

[24] Geza Kovacs, Zhengxuan Wu, and Michael S. Bernstein. 2018. Rotating Online Behavior Change Interventions Increases Effectiveness But Also Increases Attrition. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW, Article 95 (Nov. 2018), 25 pages. https://doi.org/10.1145/3274364

[25] Min Kwon, Joon-Yeop Lee, Wang-Youn Won, Jae-Woo Park, Jung-Ah Min, Changtae Hahn, Xinyu Gu, Ji-Hye Choi, and Dai-Jin Kim. 2013. Development and Validation of a Smartphone Addiction Scale (SAS). *PLOS ONE* 8, 2 (02 2013), 1–7. https://doi.org/10.1371/journal.pone.0056936

[26] Leodoro J Labrague. 2014. Facebook Use and Adolescents' Emotional States of Depression, Anxiety, and Stress. *Health Science Journal* 8, 1 (2014), 80–89.

[27] Simone Lanette, Phoebe K. Chua, Gillian Hayes, and Melissa Mazmanian. 2018. How Much is "Too Much"?: The Role of a Smartphone Addiction Narrative in Individuals' Experience of Use. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW, Article 101 (Nov. 2018), 22 pages. https://doi.org/10.1145/3274370

[28] Uichin Lee, Joonwon Lee, Minsam Ko, Changhun Lee, Yuhwan Kim, Subin Yang, Koji Yatani, Gahgene Gweon, Kyong-Mee Chung, and Junehwa Song. 2014. Hooked on Smartphones: An Exploratory Study on Smartphone Overuse among College Students *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 2327–2336. https://doi.org/10.1145/2556288.2557366

[29] Kai Lukoff, Ulrik Lyngs, Karina Shirokova, Raveena Rao, Larry Tian, Himanshu Zade, Sean A. Munson, and Alexis Hiniker. 2023. SwitchTube: A Proof-of-Concept System Introducing "Adaptable Commitment Interfaces" as a Tool for Digital Wellbeing. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 197, 22 pages. https://doi.org/10.1145/3544548.3580703

[30] Kai Lukoff, Ulrik Lyngs, Himanshu Zade, J. Vera Liao, James Choi, Kaiyue Fan, Sean A. Munson, and Alexis Hiniker. 2021. How the Design of YouTube Influences User Sense of Agency. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 368, 17 pages. https://doi.org/10.1145/3411764.3445467

[31] Kai Lukoff, Cissy Yu, Julie Kientz, and Alexis Hiniker. 2018. What Makes Smartphone Use Meaningful or Meaningless? *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 1, Article 22 (mar 2018), 26 pages. https://doi.org/10.1145/3191754

[32] Ulrik Lyngs, Kai Lukoff, Petr Slovak, Reuben Binns, Adam Slack, Michael Inzlicht, Max Van Kleek, and Nigel Shadbolt. 2019. Self-Control in Cyberspace: Applying Dual Systems Theory to a Review of Digital Self-Control Tools. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–18. https://doi.org/10.1145/3290605.3300361

[33] Ulrik Lyngs, Kai Lukoff, Petr Slovak, William Seymour, Helena Webb, Marina Jirotka, Jun Zhao, Max Van Kleek, and Nigel Shadbolt. 2020. 'I Just Want to Hack Myself to Not Get Distracted': Evaluating Design Interventions for Self-Control on Facebook. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3313831.3376672

[34] Thomas Mildner, Gian-Luca Savino, Philip R. Doyle, Benjamin R. Cowan, and Rainer Malaka. 2023. About Engaging and Governing Strategies: A Thematic Analysis of Dark Patterns in Social Networking Services. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 192, 15 pages. https://doi.org/10.1145/3544548.3580695

[35] Alberto Monge Roffarello and Luigi De Russis. 2019. The Race Towards Digital Wellbeing: Issues and Opportunities. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. ACM, New York, NY, USA, Article 386, 14 pages. https://doi.org/10.1145/3290605.3300616

[36] Alberto Monge Roffarello and Luigi De Russis. 2022. Achieving Digital Wellbeing Through Digital Self-Control Tools: A Systematic Review and Meta-Analysis. *ACM Trans. Comput.-Hum. Interact.* (nov 2022). https://doi.org/10.1145/3571810 Just Accepted.

[37] Alberto Monge Roffarello and Luigi De Russis. 2023. Defining and Identifying Attention Capture Damaging Patterns in Digital Interfaces. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA. Just Accepted.

[38] Aditya Kumar Purohit, Louis Barclay, and Adrian Holzer. 2020. Designing for Digital Detox: Making Social Media Less Addictive with Digital Nudges. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–9. https://doi.org/10.1145/3334480.3382810

[39] Seekrtech. 2023. Forest - Stay focused, be present. https://www.forestapp.cc/ Accessed: 2023-05-08.

[40] Richard H. Thaler and Cass R. Sunstein. 2008. *Nudge: Improving Decisions About Health, Wealth, and Happiness.* Yale Univ.Press.

[41] Jonathan A. Tran, Katie S. Yang, Katie Davis, and Alexis Hiniker. 2019. Modeling the Engagement-Disengagement Cycle of Compulsive Phone Use. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. ACM, New York, NY, USA, Article 312, 14 pages. https://doi.org/10.1145/3290605.3300542

[42] Yiran Wang and Gloria Mark. 2018. The Context of College Students' Facebook Use and Academic Performance: An Empirical Study. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3173574.3173992

[43] James Williams. 2018. *Stand Out of Our Light: Freedom and Resistance in the Attention Economy.* Cambridge University Press, United Kingdom.

[44] Kaitlin Woolley and Marissa A. Sharif. 2022. Down a Rabbit Hole: How Prior Media Consumption Shapes Subsequent Media Consumption. *Journal of Marketing Research* 59, 3 (2022), 453–471. https://doi.org/10.1177/00222437211055403

2

Auto play

How the Design of YouTube Influences User Sense of Agency

Kai Lukoff
University of Washington
kai1@uw.edu

Ulrik Lyngs
University of Oxford
ulrik.lyngs@cs.ox.ac.uk

Himanshu Zade
University of Washington
himanz@uw.edu

J. Vera Liao
University of Washington
ljzj@uw.edu

James Choi
University of Washington
jchoi408@uw.edu

Kaiyue Fan
University of Washington
kaiyuef@uw.edu

Sean A. Munson
University of Washington
smunson@uw.edu

Alexis Hiniker
University of Washington
alexisr@uw.edu

ABSTRACT

In the attention economy, video apps employ design mechanisms like autoplay that exploit psychological vulnerabilities to maximize watch time. Consequently, many people feel a lack of agency over their app use, which is linked to negative life effects such as loss of sleep. Prior design research has innovated *external mechanisms* that police multiple apps, such as lockout timers. In this work, we shift the focus to how the *internal mechanisms* of an app can support user agency, taking the popular YouTube mobile app as a test case. From a survey of 120 U.S. users, we find that autoplay and recommendations primarily undermine sense of agency, while
playlists and search support it. From 13 co-design sessions
that when users have a specific intention for how they wa
YouTube they prefer interfaces that support greater ag
discuss implications for how designers can help users r
sense of agency over their media use.

CCS CONCEPTS

• **Human-centered computing → Empirical studies i**

KEYWORDS

digital wellbeing, sense of agency, social media, YouTube

ACM Reference Format:
Kai Lukoff, Ulrik Lyngs, Himanshu Zade, J. Vera Liao, James Cho
Fan, Sean A. Munson, and Alexis Hiniker. 2021. How the Design of
Influences User Sense of Agency. In *CHI Conference on Human*
Computing Systems (CHI '21), May 8–13, 2021, Yokohama, Japan. A
York, NY, USA, 17 pages. https://doi.org/10.1145/3411764.344546?

CHI '21, May 8–13, 2021, Yokohama, Japan

ACM ISBN 978-1-4503-8096-6/21/05...$15.00
https://doi.org/10.1145/3411764.3445467

1 INTRODUCTION

> *"At Netflix, we are competing for our customers' time, so our competitors include Snapchat, YouTube, sleep, etc."*
>
> - Reed Hastings, Netflix CEO [117, p.50]

In the attention economy, social media apps employ a variety of design mechanisms–such as eye-catching notification icons, juicy clickbait, and never-ending autoplay–to maximize their share of
tech industry insid-
ploit psychological
ser [17, 62].

Autoplay

is often associated
port that their desire
r plans or goals and
the time [32]. And
tures of problematic

ve innovated what
manage or monitor
productivity dash-
ross many different
isms within an app.
blematic in the first

mechanisms for a
erience of being the
w sense of agency
e life impacts such
as a loss of social opportunities, productivity, and sleep [19] that often motivate digital wellbeing efforts to begin with. Moreover, a lack of sense of agency itself can be understood as a driver of the dissatisfaction that people often feel with their social media use [71].

In this work, we take the mobile app for YouTube, the most widely used social media service in the United States [90], as a test case to understand and redesign how internal mechanisms influence sense of agency. The design of YouTube must balance the interests of many different stakeholders. For example, policymakers may wish to exert control over extremist content. Advertisers may

How the Design of YouTube Influences User Sense of Agency

Kai Lukoff
University of Washington
kai1@uw.edu

Ulrik Lyngs
University of Oxford
ulrik.lyngs@cs.ox.ac.uk

Himanshu Zade
University of Washington
himanz@uw.edu

J. Vera Liao
University of Washington
ljzj@uw.edu

James Choi
University of Washington
jchoi408@uw.edu

Kaiyue Fan
University of Washington
kaiyuef@uw.edu

Sean A. Munson
University of Washington
smunson@uw.edu

Alexis Hiniker
University of Washington
alexisr@uw.edu

ABSTRACT

In the attention economy, video apps employ design mechanisms like autoplay that exploit psychological vulnerabilities to maximize watch time. Consequently, many people feel a lack of agency over their app use, which is linked to negative life effects such as loss of sleep. Prior design research has innovated *external mechanisms* that police multiple apps, such as lockout timers. In this work, we shift the focus to how the *internal mechanisms* of an app can support user agency, taking the popular YouTube mobile app as a test case. From a survey of 120 U.S. users, we find that autoplay and recommendations primarily undermine sense of agency, while playlists and search support it. From 13 co-design sessions, we find that when users have a specific intention for how they want to use YouTube they prefer interfaces that support greater agency. We discuss implications for how designers can help users reclaim a sense of agency over their media use.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI.**

KEYWORDS

digital wellbeing, sense of agency, social media, YouTube

ACM Reference Format:
Kai Lukoff, Ulrik Lyngs, Himanshu Zade, J. Vera Liao, James Choi, Kaiyue Fan, Sean A. Munson, and Alexis Hiniker. 2021. How the Design of YouTube Influences User Sense of Agency. In *CHI Conference on Human Factors in Computing Systems (CHI '21), May 8–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3411764.3445467

CHI '21, May 8–13, 2021, Yokohama, Japan

ACM ISBN 978-1-4503-8096-6/21/05...$15.00
https://doi.org/10.1145/3411764.3445467

1 INTRODUCTION

> *"At Netflix, we are competing for our customers' time, so our competitors include Snapchat, YouTube, sleep, etc."*
>
> - Reed Hastings, Netflix CEO [117, p.50]

In the attention economy, social media apps employ a variety of design mechanisms–such as eye-catching notification icons, juicy clickbait, and never-ending autoplay–to maximize their share of the user's time. In this pursuit, designers and tech industry insiders warn that many of these mechanisms exploit psychological vulnerabilities and harm the interests of the user [17, 62].

It is no accident then that social media use is often associated with a loss of sense of agency [9]. People self-report that their desire to consume media frequently conflicts with their plans or goals and that they fail to resist about three-quarters of the time [32]. And loss of control is a key component of many measures of problematic technology use [28].

In response, digital wellbeing researchers have innovated what we term *external mechanisms* that help users manage or monitor their app use, such as lockout timers [55] and productivity dashboards [57]. While these mechanisms work across many different apps, they do not change the *internal mechanisms* within an app, such as autoplay, that might lead it to be problematic in the first place.

One promising approach is to redesign these mechanisms for a greater sense of agency, i.e., an individual's experience of being the initiator of their actions in the world [112]. Low sense of agency over technology use is associated with negative life impacts such as a loss of social opportunities, productivity, and sleep [19] that often motivate digital wellbeing efforts to begin with. Moreover, a lack of sense of agency itself can be understood as a driver of the dissatisfaction that people often feel with their social media use [71].

In this work, we take the mobile app for YouTube, the most widely used social media service in the United States [90], as a test case to understand and redesign how internal mechanisms influence sense of agency. The design of YouTube must balance the interests of many different stakeholders. For example, policymakers may wish to exert control over extremist content. Advertisers may

wish to control how much time users spend on ads. Designers may wish to control how much time users spend in the app. Content creators may wish to control how much time users spend on their channel. All of these stakeholders merit consideration, however, in this work we focus specifically on *users* and how design changes might affect the control they feel over the time they spend in the mobile app.

We investigate two research questions in two studies that build upon each other:

- **RQ1: What existing mechanisms in the YouTube mobile app influence sense of agency?**
  In a survey, we asked 120 YouTube users which mechanisms make them feel most and least in control of how they spend their time in the YouTube mobile app.
- **RQ2: What changes to these mechanisms might increase sense of agency?**
  Based on the responses to the survey, we redesigned four internal mechanisms to change user sense of agency in the YouTube app: recommendations, playlists, search, and autoplay. In co-design sessions, we then asked 13 YouTube users to sketch changes of their own and evaluate our mockups. We also asked how much control they would prefer to have in different situations.

The two contributions of this work are:

(1) We identify the internal design mechanisms that influence users' sense of agency over how they spend time in the YouTube mobile app and how they might be changed. While some of these mechanisms are expected (e.g., autoplay), others are less so (e.g., playlists) and suggest promising directions for digital wellbeing (e.g., designing to support 'microplans' that guide behavior within a single session of use).
(2) We distinguish when designing for a sense of agency is desirable from when it might actually go against what users want. Participants in our co-design sessions often preferred design mechanisms that provide more control than the current version of the YouTube mobile app. This preference was stronger when they had a specific intention for using the app (e.g., to cook a recipe) than when they had a non-specific intention (e.g., to relax), in which case they still wanted to be able to turn control over to the app. We propose ways that designers might navigate this mixed preference for different levels of control at different times.

2 BACKGROUND AND MOTIVATION

2.1 Designing to Undermine Sense of Agency

Design practitioners have raised concerns about dark patterns, interfaces that are designed to manipulate a user into behavior that goes against their best interests [42, 66]. Brignull's original types of dark patterns focus on financial and privacy harms to the user [15]. However, given that people routinely report using technology in ways that are a waste of their time and that they later regret [3, 48, 58, 67], there is a need for research to examine which design patterns prompt such *attentional harms* for the user. We might term these attention capture dark patterns, designs that manipulate the user into spending time and attention in an app against their best interests.

Tech industry insiders, like the ex-President of Facebook, warn that social media apps are especially likely to innovate and employ such designs to *"consume as much of your time and conscious attention as possible"* [87]. For social games, one such a proposed pattern is "playing by appointment," wherein a game must be played according to a schedule defined by the game and not the player, or else their resources wither away [120]. Popular self-help guides promising to teach the reader how to take control of technology use (e.g., [1, 54]) suggest notifications and autoplay as design patterns to be managed or avoided. However, it is not yet established that these mechanisms necessarily lead users to feel a loss of control. For example, notifications have also been reported to reduce checking habits, since users know they will receive an alert when their desired content is ready [86].

YouTube is an important case for better understanding the design mechanisms of attention capture. YouTube has over two billion monthly users worldwide [119] and is extremely popular in the U.S., where about three-quarters of adults report using YouTube on their smartphone, with 32% using it several times a day, 19% about once per day, and 49% less often [90]. It also uses many design mechanisms that may undermine the users' sense of agency over how they spend their time on media use, e.g., notifications, autoplay, and recommendations. In particular, algorithmic recommendations drive more than 70% of YouTube watchtime [107], and a random walk analysis of YouTube recommendations found that new video recommendations get progressively longer as a user continues watching [116].

2.2 Designing to Support Sense of Agency

Reducing time spent in certain apps is a common measure of success in digital wellbeing tools. The two most popular mobile operating systems, Android and iOS, both come with default tools that let the user track and limit their time in mobile apps. Within the YouTube app itself, there are two features to manage time spent: 'Time watched statistics,' which shows how much time a user has spent on YouTube in each of the last 7 days, and the 'Take a break reminder,' which periodically prompts the user to take a rest. A strength of addressing digital wellbeing via such screentime tools is that time spent is easy to track and easy to understand.

However, a weakness of this metric is that reducing screentime is often a poor proxy for what users actually want. Instead, user intentions are often highly specific, such as wanting to reduce the time spent on targeted features of an app (e.g., on the Facebook newsfeed, but not in Facebook groups) or in certain contexts (e.g., when with family, but not when commuting on the bus) [48, 67, 70].

Within YouTube, there are two digital wellbeing features that do move beyond the time spent controls and provide more granular control. The 'notifications digest' lets a user bundle push notifications together into a single notification each day, which may reduce the triggers that lead to non-conscious, habitual use [68]. 'Autoplay toggle' lets a user decide to stop the next video from playing automatically; this may preserve the natural stopping point that comes at the end of the video, a mechanism that has been shown to help users set more deliberate boundaries around use [47]. While

the notification digest and the autoplay toggle do more than just track and limit time, it is not immediately clear by what measure of success they might be evaluated.

One promising alternative to the time spent paradigm is to design for *sense of agency*, the focus of this paper. Sense of agency is a construct that refers to an individual's experience of being the initiator of their actions in the world [112]. Sense of agency can be broken down into *feelings of agency*, that is, the in-the-moment perception of control, and *judgments of agency*, that is, the post hoc, explicit attribution of an action to the self or other [112]. In the present paper, we focus on the latter, judgments of agency.

Sense of agency matters for digital wellbeing in at least three ways. First, supporting user control is a common principle in HCI design guidelines [29, 82, 102]. Designing for an "*internal locus of control*" is one of Shneiderman and Plaisant's Eight Golden Rules of Interface Design, arising from the observation that users want "*the sense that they are in charge of an interface and that the interface responds to their actions*" [102]. Second, a low sense of control over technology use predicts greater negative life effects, e.g., internet use leading to missed social activities [19] and smartphone use leading to a risk of a loss of a significant relationship, job, or career opportunity [53]. Scales of problematic technology use generally measure both (a) lack of control and (b) negative life impacts, suggesting that 'the problem' is a combination of these two factors [20, 23]. Third, and perhaps most importantly, sense of agency matters in its own right. Feeling in control of one's actions is integral to autonomy, one of the three basic human needs outlined in self-determination theory [95]. More specific to technology use, it is also central to user (dis)satisfaction with smartphones [30, 45] and Facebook use [23, 71]. Baumer et al. note that the desire to increase *sense of agency* can even explain the seeming paradox of why users sometimes take actions that appear to reduce their *actual agency* (e.g., downloading a tool to block internet use) [9].

Prior work has investigated different means that interface designers might support sense of agency. First, designers might explore different input modalities, as, for example, keyboard input has been found to support a greater sense of agency than voice commands [64]. Second, a system's feedback should match a user's predicted feedback [63]. Third, a study of flight navigation systems found that increasing the level of automation reduced sense of agency [10]. These lessons might be revisited in the domain of digital wellbeing, as how an interface modulates sense of agency may vary with context [63].

2.3 Design Mechanisms for Digital Wellbeing

The mechanisms[1] of digital wellbeing interventions can be placed along a spectrum (see **Figure 1**). At one end are external mechanisms that monitor or police apps, such as screentime statistics and lockout timers. A hallmark of an external mechanism is that it functions identically across multiple apps, as in a timer that locks the user out of social media, gaming, *and* video apps. However, external mechanisms do not significantly change the experience within individual apps.

At the other end of the spectrum, internal mechanisms contribute to the redesign or rebuild of an experience. For example, Focus Mode in Microsoft Word redesigns the writing process by hiding all formatting options [5]. Going a step further, the standalone app *Flowstate* not only offers a minimal interface, but also deletes all text on the page if the user stops writing for longer than seven seconds [110]. Internal mechanisms fundamentally change the experience within a problematic app, or rebuild it into a new experience entirely.

At present, design researchers have innovated many tools on the external side of the spectrum, that monitor and police multiple apps at once [26, 55, 56, 78, 85]. Likewise, industry designers have built tools that apply the same time lockout mechanism to all apps, with the most popular being Digital Wellbeing for Android phones and Screentime for iPhones.

In contrast to external mechanisms, the space of internal mechanisms is relatively underexplored (see [44, 65] for notable exceptions), but holds particular promise for increasing user agency in two ways. First, designers can craft more targeted interventions with internal mechanisms than with external ones. External mechanisms, such as locking the user out of a device, often require sacrifices that users are reluctant to accept [55, 114]. Whereas an external mechanism might block the Facebook app after time is up, a more internal could configure the newsfeed to show only content from close personal friends. A redesign of internal mechanisms may be able to remove problematic aspects from an app, while still retaining its benefits.

Second, internal mechanisms shift the focus from fighting distractions to aligning interests. External mechanisms often respond to the temptations of problematic apps with microboundaries [28] or restraints on interactions [88]. However, this sets up an arms race in which the designers of digital wellbeing tools are always in a defensive position. An alternative is for designers to reenvision the internal mechanisms that lead to compulsive use in the first place [114]. Looking at the mechanisms inside of specific apps may encourage designers to not just block existing mechanisms but to innovate better ones, such as *Flowstate's* seven seconds rule for writing. This paper presents an examination how such internal mechanisms can be redesigned to support digital wellbeing.

3 STUDY 1: SURVEY OF 120 YOUTUBE USERS

Study 1 examines how existing mechanisms in the YouTube mobile app support or undermine sense of agency (**RQ1**). We wanted to start from an understanding of user experiences with the current app before moving on to designing and evaluating potential changes in Study 2 (**RQ2**). Both studies were approved by our university's IRB.

3.1 Participants

3.1.1 Recruitment. To obtain a general sample of users of the YouTube mobile app, we recruited from Amazon Mechanical Turk workers in the United States. Participants were invited to "*Help us understand how people spend their time on the YouTube mobile app.*" They were required to meet four inclusion criteria:

[1] We use the term "mechanism" to describe one component of a larger design (although some digital wellbeing designs do consist of a single mechanism)

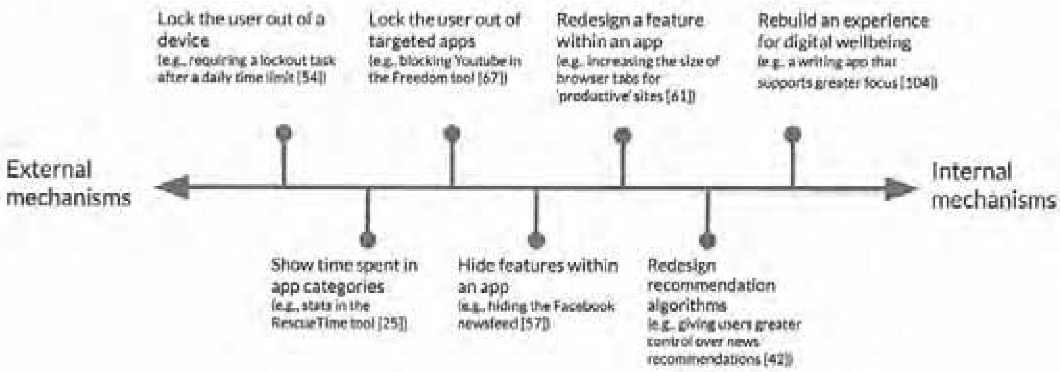


Figure 1: Mechanisms that influence how people spend their time in apps can be placed along a spectrum, as in these examples. External mechanisms monitor or police apps, while internal mechanisms redesign or rebuild the experience within a problematic app. Internal mechanisms offer designers a more targeted way of supporting user agency.

Gender identity	Man (63%), Woman (36%), Prefer not to say (1%)
Age range	18-24 (8%), 25-34 (41%), 35-44 (40%), 45-54 (11%), 55+ (1%)
Education	High school (22%), Associate degree (22%), Bachelor's degree (46%), Advanced degree (11%)
Household income (US)	<25K (14%), 25-50K (23%), 50-75K (30%), 75-125K (20%), > 125K (11%), prefer not to say (2%)
Race (choose one or more)	White (69%), Asian (17%), Black (9%), Hispanic/Latino (4%), Native American (2%)

Table 1: Demographics of the 120 survey participants

(1) A task approval rating greater than 98% for their prior work on Mechanical Turk, indicating a history of high-quality responses.
(2) Own a smartphone. Three members of our research team tested the YouTube mobile app on both Android and iPhone and found that the app has nearly identical features and only minor stylistic differences, so we accepted users of both types of devices as participants (80 Android, 40 iPhone users).
(3) Spend a minimum of 3 hours on YouTube in the past week (across all devices), according to their time watched statistics in the YouTube app. In the survey, participants saw instructions with screenshots that showed where to find this statistic in the app, confirmed that they had found it, and then entered it into the survey. To see time watched statistics, users must be signed into the app.
(4) Of the time they spend on YouTube, 20% or more is on their smartphone (self-estimated).

3.1.2 Demographics. A total of **120 participants** met the inclusion criteria and completed the survey (see demographics in **Table 1**). We excluded responses from an additional 7 participants who started but did not complete the survey. We oversampled men, Asians, and young people relative to the 2019 estimates of the United States Census Bureau [115]. Other participant samples may use the YouTube mobile app differently, e.g., users in emerging countries for whom a smartphone is often their only device for watching videos [104]. Further research is required to determine whether our results apply to other populations.

3.1.3 YouTube use. Participants spent a median of 101 minutes per day (interquartile range: 57-156) on YouTube across all devices in the week prior to the survey. Of this time, participants estimated they spent a median of 50% (interquartile range: 30-75%) in the mobile app. The median participant in our study spent more time on YouTube than the average YouTube user. In 2017, YouTube shared that signed-in users spend an average of more than 60 minutes per day in the mobile app [73] and their press page currently states that mobile accounts for over 70% of watch time [119]. We neglected to ask whether participants were using the paid YouTube premium service, which removes ads and can play videos offline and in the background; however, Google reports that only 1% of YouTube's monthly visitors subscribe to this service [108].

3.2 Procedure

People answered questions in an online survey. The initial questions asked about our four inclusion criteria. Eligible participants then continued on to background questions about their demographics and YouTube use. The complete survey wording, along with all of the other appendices for this study can be found at:https://osf.io/w3hmd

To investigate **RQ1**, one question table asked about things that made participants feel *most in control* of how they spend their time on YouTube (See **Table 2**). A second question table asked about things that made them feel *less in control*. The order of these two question tables was randomized. In terms of wording, we chose to ask about feeling "in control," as this is how sense of agency has been measured in previous studies of sense of agency in HCI (e.g., [77]) and on a self-report scale [113]. We used the informal term "*things*" because, in piloting the survey, we found that testers were unsure about whether certain things (e.g., recommendations and ads) counted as "mechanisms" of the app and we did not want to provide examples that would bias responses. In total, each participant was required to submit 6 responses for things that influenced their sense of agency on YouTube (3 for most in control, 3 for least in control).

Participants were compensated $6.00 for answering all questions, an amount that exceeds the U.S. minimum wage ($7.25 per hour).

	Thing Question: What are 3 things about the mobile app that lead you to feel most in control over how you spend your time on YouTube?	Explain Question: How does this thing make you feel more in control of how you spend your time on YouTube?
Thing 1	*"I am able to quickly access my subscribed channels."*	*"I don't spend uncontrolled amounts of time browsing through videos that may or may not be related to what I want to watch."*
Thing 2	*"I am able to get notifications of certain channels or videos getting posted."*	*"I will know exactly when a new video goes up that I may be interested in watching. This way I am not randomly checking for uploads and spending extra time searching and browsing."*
Thing 3	"Screen/watch time"	*"I can follow trends and tell when I am spending more time than usual on the app."*

Table 2: The wording and format of the "more in control" question in the survey. The example responses here come from a single study participant. All participants also completed a second version of this question table, with the text modified from "most" to "least" in the Thing Question and from "more" to "less" in the Explain Question.

The survey took a median of 21 minutes to complete (interquartile range: 15-29).

3.3 Coding reliability thematic analysis

We conducted a coding reliability thematic analysis [11, 14], in which we first established reliable codes for design mechanisms and then used them to generate themes that captured shared meanings. We started by iteratively coding the 720 responses (6 per participant). Each "thing" was analyzed as a single response, combining answers to the Thing Question and the Explain Question (i.e., one row in Table 2). In our first pass, two researchers individually reviewed all responses and met to develop initial codes. At this stage, we eliminated 112 responses without any substantive content, e.g., *"I can't think of anything else."* Of the 112 responses without substance, 55 came from "less in control" and 57 from "more."

We further limited coding to responses that specified a mechanism within the interface of the YouTube mobile app, i.e., something the app's designers could directly change. This included responses such as, *"Recommended videos - Being shown recommended videos is like a moth to a light for me,"* which was coded as 'recommendations'. It excluded responses about situational factors that are largely outside of the control of the designer such as, *"I make my own decisions - I am a conscious person who can make decisions on what I do."* This eliminated 141 more responses (59 from "less in control" and 82 from "more in control"). Interestingly, "more in control" included 28 responses that we coded as willpower, e.g., "I make my own decisions" (with only 1 such response for "less"), which suggests a potential self-serving bias [59] wherein in-control behavior is attributed to one's own willpower whereas out-of-control behavior is attributed to external factors. The other responses that we excluded were about characteristics of mobile phones (e.g., "The app is easy to access and tempt me on my phone...") and usability issues (e.g., "it crashes on me every other day or so" and "it consumes a lot of battery life") that are not specific to the interface of the YouTube mobile app. After excluding these responses, we continued with coding the 467 responses that referenced a specific design mechanism.

In our second pass, we applied the initial codes to 120 randomly selected responses and met to discuss. Since one mechanism (recommendations) came up more often than all others, we developed three subcodes for how recommendations affected participant experiences on YouTube. After merging similar codes, our codebook consisted of 21 design mechanisms, such as autoplay, playlists, and multiple device sync. In our third pass, we each independently coded the same 50 randomly selected responses. Interrater reliability was assessed using Cohen's kappa, with $\kappa = 0.73$ indicating substantial agreement [60]. In our fourth pass, we each coded half of the remaining responses, discussed the final counts, and selected several representative quotes for each code. The first author then wrote up a draft of the coding results and reviewed together with the other authors. We mapped codes (design mechanisms) to potential themes, generating three higher-level themes that structured our final writeup. In our analysis and writeup, we noted cases where responses for an individual code were split with regards to a theme, e.g., 'notifications' sometimes supported and sometimes undermined 'planning ahead'.

3.4 Results and Analysis

3.4.1 Design Mechanisms. 467 responses referenced a specific design mechanism (246 for less in control, 221 for more in control). Nine mechanisms were described as influencing sense of agency 15 or more times and are the focus of our analysis.[2] **Figure 2** provides a glanceable view of how many times each of these nine mechanisms was mentioned as leading to more or less control. **Table 3** shows the same data with a description and example response for each mechanism. **Appendix I** contains annotated screenshots that show the exact implementation of these nine mechanisms in the YouTube mobile app as they appeared when participants provided their feedback.

[2] Mechanisms mentioned 15 or more times covered 392 of 467 responses (84%) that referenced a design mechanism. Mechanisms mentioned fewer than 15 times included content moderation (13), playing videos in the background (12 responses), syncing across multiple devices (9), comments (9), ratings (8), and YouTube's 'Take a break reminder' (5). The 4 remaining mechanisms were mentioned fewer than 5 times each.

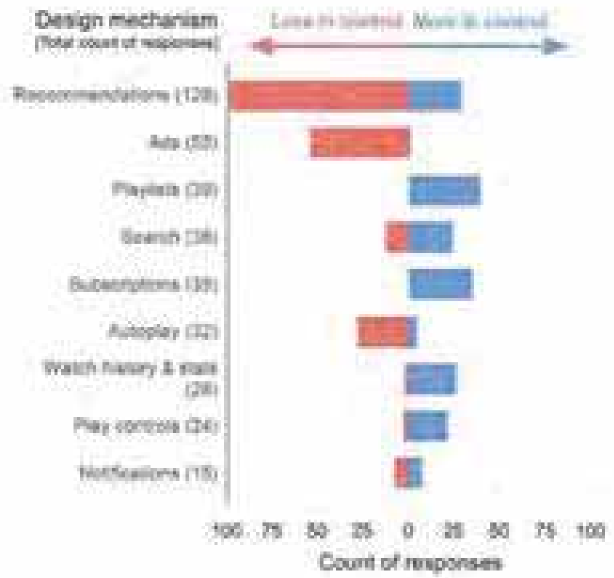


Figure 2: This diverging bar chart shows how many times these nine design mechanisms led participants to feel more control or less control. Recommendations, ads, and autoplay primarily made respondents feel less in control. Playlists, search, subscriptions, play controls, and watch history & stats primarily made respondents feel more in control. Notifications were sometimes mentioned as leading to more control and sometimes to less.

In summary, recommendations were the most frequently mentioned mechanism, accounting for 27% of all responses. Recommendations, ads, and autoplay primarily made respondents feel less in control. Playlists, search, subscriptions, play controls, and watch history & stats primarily made respondents feel more in control. Notifications were divided with about half of responses in each direction.

How Existing Mechanisms Influence Sense of Agency

The design mechanisms we identified in the YouTube mobile app informed three higher-level themes. First, users experience actions in the app along a *spectrum of consent*. Second, mechanisms for *planning ahead* help them feel more in control. Third, the *accuracy of YouTube algorithms* has mixed consequences for control. The writeup for each theme draws upon examples from our coding of the design mechanisms.

3.4.2 The spectrum of consent. Participants' sense of agency depended on whether it felt like they had 'agreed' to the actions of the app. Participants gave their active consent through actions such as tapping on a play control: "*I'm watching a video that's taken too long of my time, so I can just pause it and come back to it. I feel control there.*" Participants could also issue ongoing consent for the app, e.g., by subscribing to a creator: "*My subscriptions show me what I asked to see and I can choose what and when I wish to watch each video.*" At the other end of the spectrum were mechanisms like autoplay that acted without consent: "*It feels weird for the app to start acting before I've told it to do anything.*"

Non-consent was often felt as a result of (perceived) deception. For example, users disliked ads, but also expected them and indicated their reluctant consent. However, they seemed more upset when the app was unpredictable or violated expectations, as in: "*I understand the reason for the ads, but I don't get why some are 5 seconds and you can skip them while others are 60 seconds and you can't.*" Other cases where participants felt manipulated included when a "*small accidental click*" triggered an ad, when video creators were "*not upfront*" about the products they promoted, and when autoplay "*automatically*" turned on. Participants disliked when the app openly acted against their interests, but expressed stronger sentiments when they felt that the app also misled them about it.

3.4.3 Planning ahead. Participants felt more in control when they planned their consumption in advance. Playlists helped participants plan *how much* to watch (e.g., "*I can create playlists or queue videos in advance to limit what I watch to a specific list instead of endlessly searching around for what I want*"). Participants described the end of a playlist as a "*good place to stop*", in contrast to browsing recommendations, which they described as "endless." Watch Later, a default playlist on YouTube, also let participants control *when* and *where* to watch. A guitar teacher described how Watch Later empowered them to save videos on-the-go and watch them later in their music studio. Watch history & stats also supported planning by providing an awareness that participants could use to adjust their behavior: "*I can look at my watch history and see how many videos I have watched today. That puts it into perspective if I should spend time doing something else if I am spending too much time on YouTube.*" Several participants described using this awareness in conjunction with the Watch Later playlist: "*I am able to put a video in my Watch Later playlist if I think I have spent too much time on YouTube for the day.*"

By contrast, sense of agency was diminished by mechanisms that prompted and pressured participants with suggestions that were hard to decline. Autoplay and recommendations frequently led to this, as in "*I often spend more time than I meant to because there is a good related video that seems worth watching so ya know, 'Just one more' which becomes a couple hours.*" The Watch Later playlist again served as a safety valve in 'just one more' situations: "*Watch Later means I don't feel pressured into watching a recommended video from autoplay right when I see it.*"

Notifications sometimes supported planning and sometimes not. For example, they put participants on the spot: "*Based on my viewing history, the app will push me new content and I may not have the fortitude to not click to view.*" However, notifications also helped participants plan when to check the app by reducing their fear of missing out: "*With notifications I will know exactly when a new video goes up that I may be interested in watching. This way I am not randomly checking for uploads and spending extra time searching and browsing.*" This may explain why notifications were split between "more in control" and "less in control" responses (47% vs. 53%).

3.4.4 The accuracy of algorithms has mixed consequences for control. Irrelevant recommendations, i.e., those that were repetitive or unrelated to personal interests, universally undermined sense of agency: "*Seeing 'recommended' videos that have nothing to do

Design Mechanism	Description	Count of responses	Less in control (% of responses)	Representative quote(s) (2 quotes if minority opinion on direction of control >=20%)
Recommendations (see 3 subcodes below)	Recommended videos on the home, explore, & video player screens.	128	77%	See subcodes in the 3 rows below.
/ Irrelevant recommendations	Repetitive, uninteresting, or generic recommendations that the user is not interested in.	42 (of 128)	100%	*"The related videos are sometimes videos I've seen before, over and over."*
/ Relevant recommendations	Engaging or catchy recommendations that the user is interested in.	45 (of 128)	53%	*"YouTube has very good algorithms that know what I like, when I want it."* –VS.– *"I have a hard time not looking at the suggested videos that the algorithmpicks for me... I almost always justify watching just one more video."*
/ Customization settings	Settings to customize location, quantity, or content of recommendations.	41 (of 128)	83%	*"Not having control over the trending list. I feel like I'm force-fed content."*
Ads	Ads that appear before, during, and after videos in the player.	53	98%	*"I feel as if I am forced to watch ads, which can suck up a lot of time."*
Playlists (includes Watch Later)	Creating, saving, and playing a list of videos. Watch Later is a default playlist for all users. Playlists autoplay all videos on the list.	39	0%	*"I can create playlists or queue videos in advance to limit what I watch to a specific list instead of endlessly searching around for what I want."*
Search	Searching for videos.	36	33%	*"Very efficient and relevant searches."* –VS.– *"Countless videos have nothing to do with my latest search request."*
Subscriptions	Follow specific video creators.	35	0%	*"I can choose the content creators I want to follow so that I can limit my time to specific creators I enjoy the most."*
Autoplay	Automatically plays a new video after the current one. Can be toggled on/off.	32	87%	*"I feel like I have little control whenever YouTube takes it upon itself to just play whatever it feels like playing."*
Watch history & stats	A chronological record of videos watched and time watched stats in YouTube.	28	7%	*"I am able to view EVERYTHING I do in the app. I can keep an eye if I need to change behavior, what type of videos I watch, everything."*
Play controls	Controls to play/pause, seek forward/back, etc.	24	12%	*"I can start, pause and stop content streaming easily at any time."*
Notifications	System and in-app alerts with new subscription content, recommendations, etc.	15	53%	*"If I especially like a channel I can know about everything they upload as soon as they do"* –VS.– *"Notifications draw me to YouTube and create my schedule for 20-30 minutes. This creates an addiction."*

Table 3: This table shows nine design mechanisms that were mentioned 15 or more times in response to the survey question: "*What are 3 things about the mobile app that lead you to feel [most | least] in control over how you spend your time on YouTube?*" Design mechanisms are shown in the order of frequency of mention. The most frequently mentioned mechanism, recommendations, is shown with 3 subcodes. The representative quote(s) column shows one typical response for each design mechanism; both a "more in control" and a "less in control" quote are shown if the minority opinion on the direction of control was 20% or greater.

with my viewing history leads to unwanted scrolling and possibly unwanted content." Similarly, irrelevant search results undermined control because they forced participants to keep scrolling for what they wanted, e.g., "*I use specific search terms, but I still have to scan past a lot of vaguely or even unrelated stuff to find what I want.*"

For relevant recommendations, participants' control responses were divided nearly 50-50. In contrast to irrelevant recommendations, relevant ones supported control with their personalization (e.g., "*It has some very good algorithms that know what I like, when I want it*") or with suggestions that reached just beyond the users' comfort zone (e.g., "*I can expand my tastes based on my own preference*"). However, relevant recommendations sometimes undermined control by being *too engaging*, i.e., recommending videos that users watch, but that are unplanned and later regretted. This was captured in participants' use of terms like the "*wormhole*" (two mentions) and "*rabbit hole*" (five mentions), as in "*The way that videos get promoted to my home page and have appealing thumbnails–I end up clicking on them and wonder how I got to this place and why I am watching this video. I ended up going down the rabbit hole and watching the video and then others like it and so on.*" Some of these recommendations were described as "*clickbait*" (six mentions) that misled with content that did not meet expectations and sometimes also violated participants' consent (e.g., by showing "*inappropriate content*"). More often though, participants seemed to like the content, but felt that it was too much (e.g., "*At times there is no escape when I become interested in documentary after documentary*") or not

the right time (e.g., "*Some of the church videos are addicting and I keep watching them at night*").

Given their mixed experiences with recommendations, participants expressed frustration with the customization settings at their disposal (or lack thereof). Participants lacked the ability to customize the location, quantity, and content of recommendations. Having recommendations on almost every screen led to a loss of control: "*It seems like there are video recommendations everywhere. They are obviously in my home feed; they are in the explore menu; and they are under and beside and within other videos. It often takes me down the rabbit hole.*" Up next recommendations that appear below the current video (and autoplay after it finishes) were specifically mentioned seven times. The "*endless*" quantity of recommendations also made it hard to stop watching. Finally, participants also wanted to control *what* content is recommended, particularly when recommended content did not match their aspirations: "*There are cases in a particular day where I just want to watch cat videos. But I do not want my entire screen to recommend cat videos.*" Participants wanted to customize the content of recommendations more directly than just by generating a watch history: "*The only thing you can do to control the algorithm is to watch videos. But you get no say how it'll recommend new ones.*"

A minority of responses described recommendation settings that *do* support sense of agency. For instance, three participants appreciated how the settings menu (⋮) allows them to mark "Not interested" on specific videos, e.g., "*When I'm tempted but know a video is not educational I can hide it.*" In this case, the user *is* in fact interested in the sense that the video recommendation arouses their curiosity and attention. However, they must paradoxically mark it as "Not interested" in order to tell the interface to stop showing videos of this kind because they conflict with their longer-term goals. YouTube's settings also allow participants to delete videos from their watch history–which stops them from being used in personalized recommendations–but only one participant mentioned this feature. The vast majority of participants were either unaware of YouTube's existing customization settings for recommendations or found them inadequate.

4 STUDY 2: CO-DESIGN WITH YOUTUBE USERS

Study 1 identified existing mechanisms in the YouTube mobile app that influence user sense of agency (**RQ1**). In Study 2, we sought to understand how *changes* to these design mechanisms might influence sense of agency (**RQ2**). We conducted 13 study sessions with individual YouTube users that included two co-design activities: 1) sketching participant-generated changes; and 2) evaluating researcher-generated changes that were based on the results of Study 1. Consistent with a research-through-design approach [121], the aim of these activities was not to converge upon a single solution but rather to generate knowledge, i.e., what to design for a sense of agency.

4.1 Preparatory Design Work

In preparation for the evaluation co-design activity, five of the authors (KL, HZ, JVL, JC, KF), all advanced-degree students in a technology design program, created mockups of changes to mechanisms in the YouTube mobile app that we expected to impact sense of agency. To generate a wide range of possible changes, we started with a design brainstorm that generated 67 different ideas, e.g., creating a 'How-to mode' for viewing only educational content, reducing video playback speed to 50% after a daily time limit is exceeded, or making Watch Later the default action for recommendations. Ideas were reviewed as a group and favorites could be 'claimed' by one author who further refined it. This generated a total of 33 different sketches. We presented, discussed, and then scored these sketches according to three criteria: expected impact on sense of agency (based on the results of Study 1), novelty relative to existing digital wellbeing tools, and feasibility of implementation.[3] Expected effect on sense of agency was weighted twice in our scoring.

We created mockups for the seven sketches with the highest average scores. We wanted participants to evaluate a variety of potential changes to each mechanism, so we created three versions of each mockup: low, medium, and high-control. For example, the recommendations mechanism in the YouTube app was redesigned to change the number of recommendations shown on the homepage, with the low-control version showing unlimited recommendations, the medium-control version showing only three recommendations with a button to "show more," and the high-control version not showing any recommendations (see images in **Table 4**). To focus on RQ2, our results and analysis here address only the four mockups (see **Table 5**) that directly change one of the existing internal mechanisms in YouTube that we identified in Study 1. The other three mockups we tested—activity-goal setting, time-goal setting, and a timer—are more external mechanisms that might apply equally well to other apps. However, we decided to focus this paper on the unique potential of internal mechanisms.

We note that although our research focuses at the level of 'design mechanisms,' the details of these designs matter. For instance, although the recommendations in the current version of YouTube seemed to reduce sense of agency in most of the Study 1 responses, a different implementation of 'recommendations' might produce different effects. This is true of our mockups too: in our search redesign we showed a task-oriented example query ("*How to cook a turkey*"), whereas a leisure-oriented example query (e.g., "*Funny cat videos*") could have led to different results. We include descriptions of the most relevant details of each of these design mechanisms in the body of the paper, screenshots of their current implementation in the YouTube mobile app in **Appendix I**, and images of *all* of our mockups in **Appendix II**.

4.2 Participants

4.2.1 Recruitment. We recruited YouTube users in Seattle via email lists and social media channels to "*Help us understand how people spend their time in the YouTube mobile app.*" We did not initially set inclusion criteria for participation (beyond adult YouTube users) as we viewed our co-design activities as exploratory. However, after our initial sessions proved insightful for our team of design researchers, we sent a follow-up survey to participants that asked

[3] Feasibility was a criterion to focus on designs that a third-party mobile developer could build using public APIs, an intention we have for our future work.

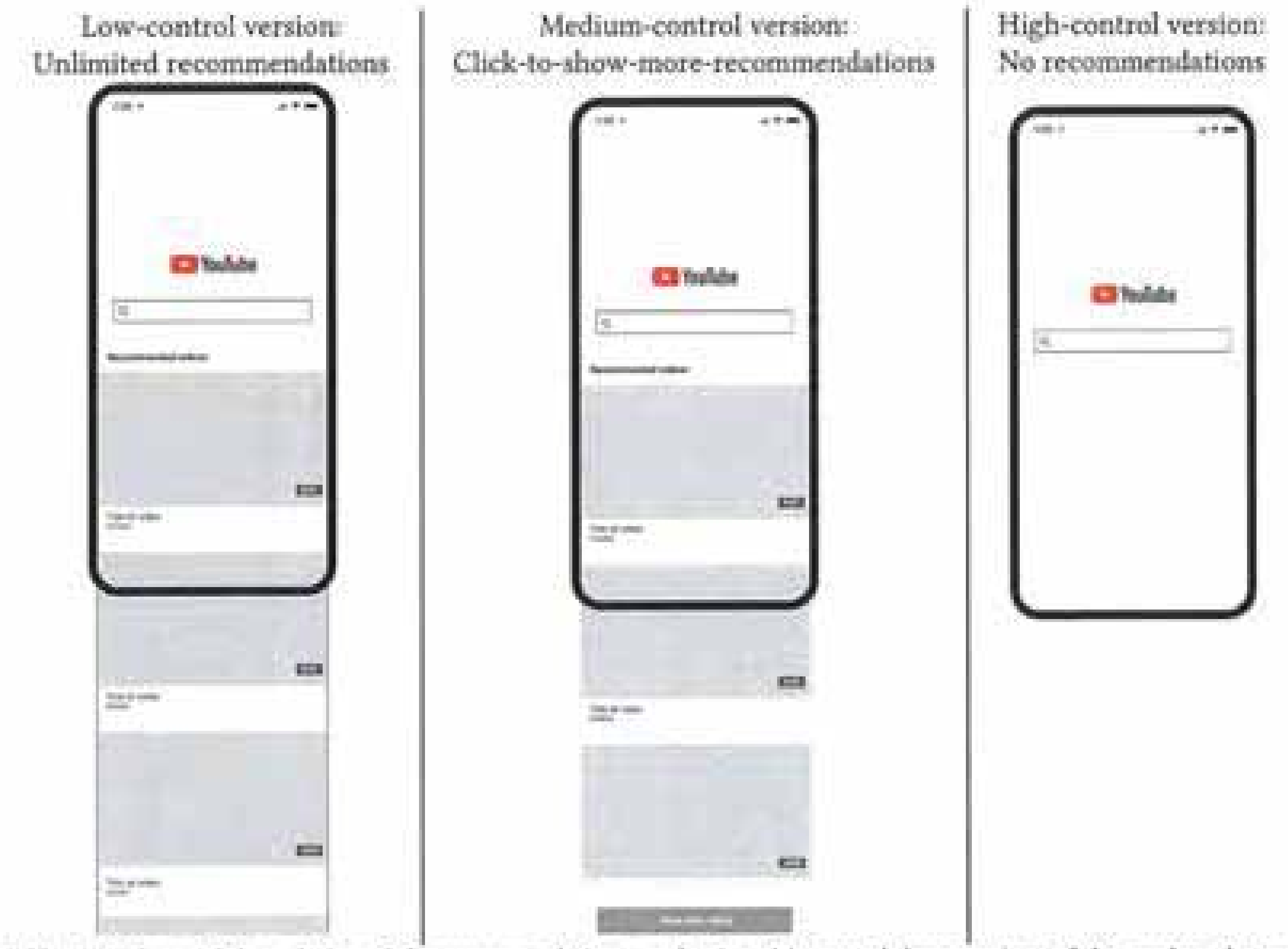


Table 4: Mockups of the redesign of the recommendations mechanism. We created three versions of the mockup that we expected to offer different levels of control. These 3 versions of each redesign were evaluated by participants in the co-design evaluation activity.

Redesigned mechanism	Dimension of change	Low-control version	Medium-control version	High-control version	Related experience for users (as described by Study 1 participants)	Comparison to current version of YouTube mobile app
Recommendations	Number of video recommendations on home screen	Unlimited recommendations	Shows 3 recommendations, then a click-to-show-more button	No recommendations	Endless recommendations often undermine sense of agency	Similar to low-control version
Playlists	Prominence of button to save a video in the Watch later playlist	No Watch Later button	Small Watch later button	Large Watch Later button	Watch Later playlist lets users plan ahead, reduces pressure to watch now	Similar to medium-control version
Search	The degree to which search prioritizes fun vs. relevant results (see **Appendix II** for more details)	Prioritize "fun" results (intended to be too engaging)	User can toggle between "fun" & "relevant" results	Prioritize "relevant" results	Sometimes recommendations and search results that are too engaging undermine sense of agency	Similar to medium-control version
Autoplay	The degree of user consent required to play the next video recommendation	Autoplay the next recommendation	Show the next recommendation	No next recommendation	Autoplaying videos without consent undermines sense of agency	Similar to low-control version

Table 5: This table describes our redesigns of 4 existing mechanisms in the YouTube app. We created three versions of each mockup that we expected to provide different levels of control to the user: low, medium, and high. Appendix II describes more details about the search redesign and the three additional mockups we created, which we do not report on here.

about demographics and YouTube use. Participants were compensated with a $30 voucher.

4.2.2 Demographics and YouTube use. 13 YouTube users (7 women, 6 men) participated in our sessions. The median age was 29 (range: 18-36). Participants reported using YouTube a median of 52 minutes per day (range: 27-70), again based on checking their time watched statistics in the YouTube mobile app. For reference, this amount of time is slightly lower than the average of signed-in YouTube users (60 minutes) [73] and considerably lower than the median of participants in Study 1 (101 minutes).

4.3 Procedures

Sessions included an initial think-aloud demonstration of their current YouTube use, followed by sketching and evaluation co-design activities. The median length of a session was 75 minutes (range: 57-105 minutes).

4.3.1 Think-aloud Demonstrations with YouTube App. In a modified version of a think-aloud-protocol [52], the participant opened YouTube on their smartphone and talked us through a typical engagement cycle (how they start and stop use) [114]. Next, they showed and talked us through the mechanisms that made them feel most and least in control of how they spend their time on YouTube.

4.3.2 Co-design Activity 1: Sketching. To elicit participant-generated ideas, we asked participants to sketch over paper mockups of three key screens: home, search, and video player (see **Figure 3**). Each screen represented a minimal version of a video app without recommendations, rather than a direct copy of the current YouTube interface. We chose this minimal version to encourage participants to generate new ideas, rather than to evaluate the existing interface (which we did in Study 1). Participants were handed a pen and a copy of one mockup (e.g., the home screen) and were asked, "*What would you change on this page to feel more in control of how you spend your time on YouTube?*" They then received a second copy of the same mockup and were asked to sketch changes that would make them feel "*less in control.*" Each participant created a total of six sketches (two versions of three different screens). As they sketched, participants were asked to explain their thinking [98].

4.3.3 Co-design Activity 2: Evaluation. To receive feedback on our changes from YouTube users, we asked participants to evaluate our mockups of the redesigned mechanisms in the YouTube mobile app (see **Table 5**). For each mockup, the three different versions were placed in front of the participant in a random order, they reviewed for about one minute, and then asked any questions they had. We did not tell participants which one was the low, medium, or high-control version. The participant was then asked to rank the three versions in order from the one they would least prefer to use to the one they would most prefer, and explain why.

4.4 Codebook Thematic Analysis

We used codebook thematic analysis to analyze the data [12, 14], wherein we generated themes that are more interpretive than just

Figure 3: Sketches of the home screen of the YouTube mobile app. The participant (P11) explained that in the "more in control" version, recommendations are based on topics chosen by the user. In the "less in control" version, the user needs to scroll through recommendations to see the search bar at the bottom of the screen.

a summary of all of the data, but less interpretive than in reflexive thematic analysis where the researcher's subject position plays a central role in the analysis [13]. After each co-design session, the researcher leading the session completed a debriefing form with their top three takeaways and shared participant sketches with the rest of the research team. We held weekly meetings to discuss these data and discuss initial ideas. After finishing data collection, all co-design sessions were transcribed. To further familiarize ourselves with the data, three of the authors read the transcripts and again reviewed the sketches. We next independently coded the data using a web app for collaborative coding [103] to generate our set of initial codes. After reviewing this first pass of coding together, we refined and consolidated codes and generated initial themes. Our final set of codes included: user freedom of choice, situational features affecting control, design mechanisms for control, setting clear expectations for the user, and triggers to stop, each of which had further subcodes. We applied our codes to all transcripts and sketches and reviewed the results to create our final themes. For each theme, we extracted vivid exhibits [6], which we used to write analytical memos.

4.5 Results and Analysis

We generated two themes about how participants expected changes to the design mechanisms of YouTube would affect their sense of agency. First, participants wanted design mechanisms that provided *more control when they had an intention in mind* as opposed to when they just wanted to explore. Second, participants envisioned and wanted mechanisms for *active and informed choices* to increase control.

4.5.1 Specific intentions call for more control. When individual participants reviewed the different versions of their own sketches and our mockups, they were often conflicted about how much control they preferred. It depended upon the situation. When they had a specific intention or goal for their YouTube visit (e.g., to cook a recipe), they wanted design mechanisms that provided greater control. When they had a non-specific intention such as relaxing, they preferred design mechanisms that turned control over to YouTube.

For participants, specific intentions varied from watching a video of a favorite dance, to the latest basketball highlight, to a tutorial on solving a Rubik's Cube. When they had such a specific intention in mind, they wanted greater control than YouTube currently gives them. P4 removed recommendations from their sketch, explaining: "*If I have a specific goal, I know what I want, I don't need recommendations to guide my search, I just want to be in control of my search.*" P2 evaluated our redesign of the search mechanism that emphasized results with higher entertainment value by saying, "*I'm probably going to click on it because it's cute and I'm just going to waste so much time. So it's going to make me feel totally out of control of what I actually wanted to come here for.*" In these cases, participants wanted stronger control mechanisms so that the app would not hijack their specific intention.

Sometimes participants held intentions with a moderate level of specificity, in which case participants wanted to retain some control but also delegate some to YouTube. Often these intentions were topical, as in when P11 wanted to be able to use the app in an "*active way*" to search and browse videos about programming, but not in a "*passive way*" to follow just any recommendation. Sometimes, these intentions were temporal, such as when working or studying, participants preferred a version of YouTube that helps them watch a moderate number of videos without making them "*fall down a rabbit hole of similar related stuff*" (P13). To address these cases, participants sketched both changes to internal mechanisms that were specific to YouTube (e.g., limits on the number of recommended videos) and also more external mechanisms that might apply across a variety of social media apps (e.g., time reminders).

By contrast, when participants had only a non-specific intention (e.g., to unwind or explore), they wanted YouTube to lead the way. Our redesigns of the recommendations mechanism showed either unlimited, limited, or no video recommendations, to which P2 responded "*If I came here for a specific reason, like my goal is to learn how-to do something, then I prefer this one without recommendations. However, if I just want to watch something that gets my mind off things, I prefer the one where I can choose to show more recommendations.*" At times when participants just wanted to be entertained, designing for greater control could actually get in the way. P13 shared, "*If you're not giving me recommendations, and if you're making me search, then I'm not in control. Or, I'm in control, but the problem is I'm spending more time. There's no point.*"

4.5.2 Active and informed choices. The Study 1 theme "Spectrum of consent" addressed whether the user had 'agreed' to an action taken by the app (e.g., autoplaying the next video). To support control, Study 2 participants envisioned more active choices, where the user felt like they were the one to initiate the action. As a step in this direction, P1 described a home screen that presented, "*Six categories we think you're most interested in, and then you're at least making the active choice, 'I want to watch some interviews right now.'*" In this design, the app's algorithm would recommend a set of personalized topics, but the user would be the one to choose between them. A still more active choice was when the user was the one to generate the set of choices in the first place, as in P7's sketch: "*There aren't a billion recommendations on the home screen. It's just a search bar. You go straight to what you want to watch, you watch it, and then you're done.*" Participants described search as a paragon of user-led choice, and many foregrounded the search option in their sketches to increase control and hid it in ones to decrease control (see **Figure 3**).

Many sketches also supported more *informed* choices. These designs made it easier for users to know what to expect from a video by surfacing metadata like view count, user ratings, and descriptions. Five participants proposed novel metadata, such as an 'activity time' filter that would sort how-to videos by the time it takes to perform the activity they teach, e.g., cook a recipe (P12). Another suggested expert ratings as an indicator of quality (P5). Conversely, in sketches to undermine control, it was common to remove video metadata. P12 likened this to the experience at Costco, a supermarket chain that deliberately shows no signs in its stores [83]: "*If you want to go find cookies, they won't actually show you where the cookies are so you literally have to go through every single aisle. You have to go find it.*"

More choice alone did not lead to more control. In sketches of designs to undermine control, participants covered every corner of the home screen with video recommendations that scrolled infinitely (P11) and in every direction (P5). P13 described, "*If they didn't have [recommended videos], it would be a lot harder to follow these different rabbit holes. I imagine that I would have to intentionally seek out another video, so I wouldn't feel sucked in as much.*" Recommendations prompted a passive form of choice, in which users reacted to the app's infinite scroll of suggestions, rather than making active choices on their own terms.

5 OVERALL DISCUSSION

Together, our two studies identify design mechanisms that influence sense of agency in the YouTube mobile app and how they might be changed to increase it. In Study 1, participants reported that, in the current app, recommendations, ads, and autoplay mostly led them to feel less in control, whereas playlists, search, subscriptions, play controls, and watch history & stats mostly made them feel more in control. Across all existing mechanisms, participants felt less in control when the app took actions of its own without their consent (e.g., autoplaying a new video recommendation). Recommendations were of special concern and participants expressed frustration at their inability to customize their location, quantity, and content. In

contrast, by helping participants plan ahead for even just a short while, existing mechanisms like playlists and watch stats made participants feel more in control.

When participants envisioned and evaluated changes in Study 2, they wanted more opportunities to make active choices, rather than respond to a set of choices proposed by the app. This preference was stronger when they had a specific intention in mind (e.g., to watch a certain video or topic), whereas when their intention was more general (e.g., to pass the time) they favored turning control over to YouTube.

We expect that our findings on how design mechanisms influence sense of agency on YouTube are most likely to generalize to other social media and media apps where users (a) report feeling out of control at times (e.g., Facebook [71]); and (b) use the app for both specific and non-specific intentions (e.g., Pinterest [23]). We first discuss our findings mostly with respect to our test case of YouTube, before considering implications for digital wellbeing more broadly.

5.1 Rethinking What 'Relevance' Means for Recommendations

Recommendations were mentioned by participants as undermining sense of agency far more times than any other design mechanism in the YouTube mobile app, suggesting that recommender systems [92] should be of central concern to digital wellbeing designers. However, they led to a reduced sense of agency via two very different routes: irrelevance and relevance.

First, recommendations were sometimes irrelevant, showing videos that participants were simply not interested in. However, due to rapid advances in artificial intelligence and recommender systems like YouTube specifically (e.g., [27]), one might expect recommendations in social media apps to become more and more relevant in the coming years.

Second, recommendations were sometimes too 'relevant,' which presents a more vexing problem from a digital wellbeing perspective. For example, participants reported that they sometimes saw *too many* interesting recommendations (e.g., for documentaries or for church videos late at night), which made them feel a loss of control. In this case, YouTube's algorithm is arguably *too good* at a local optimization problem: *Out of millions of videos, which one is the user most likely to watch?* But it misses a more global optimization problem: *Out of many possible actions, which one does the user most want to take?* In these cases, recommendations appealed to a users' impulse or short-term desire to watch more videos, but conflicted with their long-term goals, creating a self-control dilemma for the user [35, 69].

Our findings call for rethinking what 'relevance' means for recommendations in the context of digital wellbeing. Prior research on recommender systems has argued that *"being accurate is not enough,"* as a fixation on accuracy can lead designers to ignore important facets of user experience like serendipity [76, p.1]. For participants in our study, sense of agency was clearly a neglected facet of user experience, as YouTube's recommendations led them to actions (i.e., watching more videos) they did not feel they controlled. To be clear, this does not mean that Google or others should try to create an 'algorithm for life' that recommends between watching another video, writing a term paper, and going to sleep.

However, it does suggest that recommender systems could first start with the global problem of *when* to show recommendations, before moving on to the local problem of *which* items to recommend. For example, a decision *not* to show recommendations might be informed by the time of day (e.g., 2am is too late), screentime preferences (e.g., when the user has already exceeded their goal of 30-minutes per day on entertainment apps), or explicit user preferences (e.g., only show three recommendations unless I click-to-show-more). In HCI research, sometimes the implication of a user needs assessment is *not* to design technology, as a new technology might not be appropriate in the context of the larger situation [8]. Similarly, for recommender systems, our findings suggest that sometimes the implication is *not* to recommend. Prior work has addressed how a system can display the level of confidence it has in its recommendations to the user [75], but this should be preceded by a more fundamental question of whether or not to show recommendations in the first place.

Whereas both of the studies in this work elicit user preferences ("what users say"), the dominant paradigm of recommender systems today, including YouTube, is behaviorism: recommendations largely neglect explicit preferences and instead rely on behavior traces ("what users do") [36]. The present bias effect [84] predicts that *actual behavior* will favor the choice that offers immediate rewards at the expense of long-term goals. In this way, recommender systems reinforce the sometimes problematic behavior of the current self rather than helping people realize their 'aspirational self' that reflects long-term goals [36, 68].

Participants also wanted to customize the *content* of recommendations, e.g., *"I do not want my entire screen to recommend cat videos."* Today, the dominant paradigm of recommender systems, including YouTube, is behaviorism: recommendations rely on behavior traces ("what users do") and largely neglect explicit preferences ("what users say"). In this way, recommender systems reinforce the sometimes problematic behavior of the current self rather than helping people realize their 'aspirational self' that reflects long-term goals [36, 68]. Designers might address this by making it easier for users to (a) explicitly state preferences for topics they would like to see or not see; (b) explicitly rate recommendations (e.g., show me more like this one); (c) edit their viewing history to influence future recommendations (e.g., delete all cat videos); or (d) select an algorithmic personae to curate their recommendations (e.g., "The Diplomat," who brings news videos from the other side) [44, p.72]. The current YouTube app offers limited support for these first three features (e.g., users can select from among topics for recommendations on the home page of the app), but participants in our study were mostly either unaware of these customization settings or found them to be inadequate.

To summarize, we encourage designers of recommender systems to think more broadly. This includes exploring how recommendations support user aspirations rather than just reinforce current behaviors, which may require identifying different measures on which to optimize. Designers and researchers should also continue to explore features for customizing or tuning recommendations to user needs, and designing those customizations that put users in control - at least to the extent they want.

5.2 Designing to Support Microplanning

Behavior change researchers have long known that plans can help bridge the gap between intentions and behavior. In this work, plans are usually crafted in advance through careful deliberation and guide behavior for some time into the future [2]. For example, a screentime tool in this mold might ask the user to review and reflect upon their past usage data and develop a plan for their use over the next month. Participants in our study also 'planned', but they did so in a more ad hoc manner. For example, they queued videos in advance to limit what they watched during a single session or glanced at their Time watched statistics to know whether to watch another video or add it to their Watch Later playlist.

These types of actions might be called 'microplanning,' making lightweight plans that guide behavior for a short time, usually just a single session of use. Our naming takes inspiration from Cox et al.'s coining of the term 'microboundary' to describe "*a small obstacle prior to an interaction that prevents us rushing from one context to another*," which serves as a 'micro' version of a commitment device that prevents the user from "*acting hastily and regretting it later*" [28]. 'Microboundary' has helped center an important concept from behavioral economics, commitment devices that restrict future choices to reflect long-term goals [16, 96], in the research and development of digital wellbeing tools, e.g., [55, 56, 69, 91].

Similarly, we hope that the concept of 'microplans' encourages the use of behavior planning knowledge in the design of digital wellbeing tools. For example, this literature finds that plans are more likely to succeed if they specify where, when, and how a behavior will be enacted [40]. A microplan might incorporate just the 'where' part, and be supported by a video playlist that is tied to a specific location, e.g., song tutorials *for my guitar studio*. Triggers are also a key component of effective plans [38], so in this case the playlist might be the primary recommendation in the app anytime the user is within 50 meters of the studio. In another example, Hiniker et al. adapted an evidence-based Plan-Do-Review sequence [37] for an app that asked children to plan out their video-watching, finding that it helped them transition to their next activity with ease [49]. In the domain of impulse buying [80], an e-commerce site (or third-party extension) might foreground 'shopping list' tools to support intentional buying.

5.3 Different Levels of Control for Ritualized and Instrumental Use

In Study 2, participants suggested ways that the YouTube mobile app might be redesigned to increase sense of agency (e.g., by reducing the number of recommendations it displays). However, such changes might lead to adverse effects as there were also times when participants preferred low-control features. Although HCI design guidelines advise supporting user sense of agency [82, 102], we should not assume that a greater sense of agency is always desirable.

Specifically, participants preferred higher-control mechanisms when they had a specific intention in mind and lower-control ones when they had a non-specific intention. This finding broadly aligns with two types of viewing that have been identified in uses and gratifications research on television use [94]: (1) ritualized use, open-ended use to gratify diversionary needs; and (2) instrumental use, goal-directed use to gratify informational needs. On this view, the current version of the YouTube app appears to offer good support for ritualized use, but poor support for instrumental use, as participants often felt that their specific intentions were hijacked by its autoplay and endless recommendations.

How might a single app support sense of agency for both ritualized *and* instrumental use? One approach is to let the user switch between low and high-control interfaces. This can be done at the app-level, e.g., switching between an Explore Mode and a Focus Mode. Or it can be done at a feature-level, e.g., YouTube currently offers an on/off toggle for autoplay, but does not provide any way to toggle recommendations, even though study 1 survey respondents say recommendations undermine sense of agency far more often than autoplay. Power users in particular may prefer an interface that is customizable (user-tailored) by a toggle, whereas non-power users may prefer one that is personalized (system-tailored) for them [111].

A second approach is to change the user interface based on a personalized prediction model. Recent work has found that classifiers can be trained to predict these types of media use with high confidence, e.g., for Pinterest [24] and smartphone use [50]. For example, if YouTube expects that the user is visiting for ritualistic use, it could remain as is, or even go further to take control as in its Leanback mode for "*effortless viewing*" that autoplays a never-ending stream of high-definition recommendations [41]. Both our own findings on autoplay and previous work suggest that such a high level of automation would reduce sense of agency [10], but it may still be the interface that the user prefers in this situation. Conversely, if YouTube has high confidence that the user is visiting for instrumental use, it could present a search-only interface and hide all recommendations. Finally, if it has low confidence in its prediction, it could present a middle-ground interface that shows limited recommendations, or it might err on the side of caution and lead with a search-first interface in case the user has an intention to express.

5.4 Towards a Language of Attention Capture Dark Patterns

Our findings address *what* and *when* to design to increase sense of agency. However, in the attention economy, what might motivate key stakeholders to support such designs? One step is for the design community to develop a common language of attention capture dark patterns that recognizes designs that lead to attentional harms.

Developing such a lingua franca of attention capture design patterns could be integrated into design education [42], influence designer thinking, and reputations, as is done by the name-and-shame campaign of the darkpatterns.org website [15]. At the company level, it could help inspire products that are mindful of the user's sense of agency. For example, in spite of the incentives of the attention economy, Apple is now working to make *privacy* a selling point [43], e.g., by preventing developers from tracking users across multiple apps without their active consent [4]. At the regulatory level, a recent review of dark patterns by Narayanan et al. notes that if the design community does not self-regulate by setting standards for itself, it may be regulated by more onerous standards set by others [81]. The U.S. Senate is currently considering how to

regulate social media, with one bill that would make it illegal to "*manipulate a user interface with the purpose or substantial effect of obscuring, subverting, or impairing user autonomy*" [74] and another that would ban autoplay and infinite scroll [22]. For designers, the language of dark patterns is an important way to contribute to a broader critical discussion of design practices in the the technology industry [42].

We caution that the message of attention capture dark patterns should *not* be "never X," but rather "be careful when X." Participants in both of our studies reported mixed experiences with many design mechanisms, including autoplay and recommendations. An outright ban on these mechanisms is likely to reduce sense of agency in a substantial number of situations where the user just wants to explore. Instead, a nuanced guide to dark patterns might present examples of the problem, followed by counterexamples where such a pattern is appropriate. While this creates a murky gray middle, it also better describes the effects of the design mechanisms that we identified in our studies.

5.5 Limitations

In addition to the previously stated limitations of our participant sampling and focus on design mechanisms as a unit of analysis, our work also has at least four conceptual limitations that could be explored in future work. First, both of our studies asked participants to share their preferences, however present bias [84] predicts that *actual behavior* will favor the choice that offers immediate rewards at the expense of long-term goals. An *in-situ* study of how people respond to redesigns intended to influence sense of agency would yield results on ("what users do"), which might need to be reconciled with the present results on ("what users say"). Second, time and attention are not the only factors that influence sense of agency. By asking participants in both studies to reflect on "*...in control of how you spend your time on YouTube*" we discouraged participants from considering other factors such as privacy [111]. In Study 2, this may have primed participants to focus on sense of agency over other factors when evaluating which version of the mockup they preferred. Third, self-reported agency can be quite different from the facts of agency [29, 79]. For example, many people continue to press 'placebo buttons' like the 'close door button' in their apartment's elevator, even when doing so has no effect [89]. There is therefore a concern that designs to increase sense of agency may be disconnected from actual ability to influence the world. Fourth, users are not the only stakeholders on YouTube, and it would be a mistake to optimize for their sense of agency alone. Google, creators, advertisers, and even society itself all have a stake in what happens on YouTube. For instance, radicalizing political videos can make viewers feel as if they have uncovered powerful conspiracies that were previously hidden from them [93]; to support sense of agency in this use case would be dangerous. User sense of agency needs to be integrated into larger design frameworks as one important consideration among many for social media apps.

6 CONCLUSION

Whereas a common approach to digital wellbeing is designing to reduce screentime, this work takes an alternative approach of designing to increase sense of agency. In two studies, we identify mechanisms within the YouTube mobile app that participants report influence their sense of agency and how they want to change them. We find that participants generally prefer mechanisms like autoplay and recommendations to be redesigned for a greater sense of agency than the YouTube mobile app currently provides. For digital wellbeing designers, we highlight a need for recommender systems that better reflect user aspirations rather than just reinforce their current behavior. We also propose mechanisms that support 'microplanning,' making lightweight plans to guide a single session of use, to increase user sense of agency. Finally, we propose language that the design community might adopt to recognize design patterns that impose attentional harms upon the user.

ACKNOWLEDGMENTS

This work was funded in part by National Science Foundation award #1849955. We thank Xuecong Xu, Ming Yao Zheng, Kevin Kuo, Tejus Krishnan, Laura Meng, Linda Lai, and Stefania Druga for helping to conceptualize this study and design the mockups.

REFERENCES

[1] [n.d.]. Take Control. https://www.humanetech.com/take-control Accessed: 2020-8-3.

[2] Elena Agapie. 2020. *Designing for Human Supported Evidence-Based Planning*. Ph.D. Dissertation. https://digital.lib.washington.edu/researchworks/handle/1773/45709

[3] Morgan G Ames. 2013. Managing Mobile Multitasking: The Culture of iPhones on Stanford Campus. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work (San Antonio, Texas, USA) (CSCW '13)*. ACM, New York, NY, USA, 1487–1498. https://doi.org/10.1145/2441776.2441945

[4] Apple Inc. [n.d.]. Details for app privacy questions now available - News - Apple Developer. https://developer.apple.com/news/?id=hx9s63c5 Accessed: 2020-9-15.

[5] Catherine Baab-Muguira. 2017. The Stupidly Simple Productivity Hack Hiding In Microsoft Word. https://www.fastcompany.com/3068825/the-stupidly-simple-productivity-hack-hiding-in-microsoft-word Accessed: 2020-9-11.

[6] Liam Bannon, John Bowers, Peter Carstensen, John A Hughes, K Kuutti, James Pycock, Tom Rodden, Kjeld Schmidt, Dan Shapiro, Wes Sharrock, and Others. 1994. *Informing CSCW system requirements*. Lancaster University. https://www.forskningsdatabasen.dk/en/catalog/2185760093

[7] Eric P S Baumer, Phil Adams, Vera D Khovanskaya, Tony C Liao, Madeline E Smith, Victoria Schwanda Sosik, and Kaiton Williams. 2013. Limiting, Leaving, and (Re)Lapsing: An Exploration of Facebook Non-use Practices and Experiences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Paris, France) (CHI '13)*. ACM, New York, NY, USA, 3257–3266. https://doi.org/10.1145/2470654.2466446

[8] Eric P S Baumer and M Six Silberman. 2011. When the implication is not to design (technology). In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2271–2274. https://doi.org/10.1145/1978942.1979275

[9] Eric P S Baumer, Rui Sun, and Peter Schaedler. 2018. Departing and Returning: Sense of Agency As an Organizing Concept for Understanding Social Media Non/Use Transitions. *Proc. ACM Hum. -Comput. Interact.* 2, CSCW (Nov. 2018), 23:1–23:19. https://doi.org/10.1145/3274292

[10] Bruno Berberian, Jean-Christophe Sarrazin, Patrick Le Blaye, and Patrick Haggard. 2012. Automation technology and sense of control: a window on human agency. *PloS one* 7, 3 (March 2012), e34075. https://doi.org/10.1371/journal.pone.0034075

[11] Richard E Boyatzis. 1998. *Transforming Qualitative Information: Thematic Analysis and Code Development*. SAGE. https://play.google.com/store/books/details?id=_rfClWRhIKAC

[12] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (Jan. 2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[13] Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health* 11, 4 (Aug. 2019), 589–597. https://doi.org/10.1080/2159676X.2019.1628806

[14] Virginia Braun, Victoria Clarke, Nikki Hayfield, and Gareth Terry. 2018. Thematic Analysis. In *Handbook of Research Methods in Health Social Sciences*, Pranee Liamputtong (Ed.). Springer Singapore, Singapore, 1–18. https://doi.org/10.1007/978-981-10-2779-6_103-1

[15] Harry Brignull and Alexander Darlington. [n.d.]. What are dark patterns? https://www.darkpatterns.org/ Accessed: 2019-9-28.

[16] Gharad Bryan, Dean Karlan, and Scott Nelson. 2010. Commitment Devices. *Annual review of economics* 2, 1 (Sept. 2010), 671–698. https://doi.org/10.1146/annurev.economics.102308.124324

[17] Christopher Burr, Nello Cristianini, and James Ladyman. 2018. An Analysis of the Interaction Between Intelligent Software Agents and Human Users. *Minds and Machines* 28, 4 (Sept. 2018), 735–774. https://doi.org/10.1007/s11023-018-9479-0

[18] calkuta. [n.d.]. DF Tube (Distraction Free for YouTube). https://chrome.google.com/webstore/detail/df-tube-distraction-free/mjdepdfccjgcndkmemponafgioodelna?hl=en Accessed: 2020-8-5.

[19] Scott E Caplan. 2010. Theory and measurement of generalized problematic Internet use: A two-step approach. *Computers in human behavior* 26, 5 (Sept. 2010), 1089–1097. https://doi.org/10.1016/j.chb.2010.03.012

[20] Hilarie Cash, Cosette D Rae, Ann H Steel, and Alexander Winkler. 2012. Internet Addiction: A Brief Summary of Research and Practice. *Current psychiatry reviews* 8, 4 (Nov. 2012), 292–298. https://doi.org/10.2174/157340012803520513

[21] Marta E Cecchinato, Anna L Cox, and Jon Bird. 2017. Always On(line)?: User Experience of Smartwatches and their Role within Multi-Device Ecologies. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 3557–3568. https://doi.org/10.1145/3025453.3025538

[22] Angela Chen. [n.d.]. A new bill would ban making social media too addictive. *MIT Technology Review* ([n. d.]). https://www.technologyreview.com/2019/07/30/133976/josh-hawley-social-media-addictive-design-legislation-smart-act-bill/

[23] Justin Cheng, Moira Burke, and Elena Goetz Davis. 2019. Understanding Perceptions of Problematic Facebook Use: When People Experience Negative Life Impact and a Lack of Control. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 199. https://doi.org/10.1145/3290605.3300429

[24] Justin Cheng, Caroline Lo, and Jure Leskovec. 2017. Predicting Intent Using Activity Logs: How Goal Specificity and Temporal Range Affect User Behavior. In *Proceedings of the 26th International Conference on World Wide Web Companion*. International World Wide Web Conferences Steering Committee, 593–601. https://doi.org/10.1145/3041021.3054198

[25] Julie E Cohen. 2012. What privacy is for. *Harvard law review* 126 (2012), 1904. https://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/hlr126&section=88&casa_token=WL8sq3iJD9YAAAAA:ET2vm5L-cBu99C528GgMfSdFRCGxFo-aMK6PlepmiwFmC6VWTYJsbwQT5AXRYoi42VexRjoT

[26] Emily I M Collins, Anna L Cox, Jon Bird, and Cassie Cornish-Tresstail. 2014. Barriers to Engagement with a Personal Informatics Productivity Tool. In *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: The Future of Design* (Sydney, New South Wales, Australia) *(OzCHI '14)*. ACM, New York, NY, USA, 370–379. https://doi.org/10.1145/2686612.2686668

[27] Paul Covington, Jay Adams, and Emre Sargin. 2016. Deep Neural Networks for YouTube Recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems* (Boston, Massachusetts, USA) *(RecSys '16)*. Association for Computing Machinery, New York, NY, USA, 191–198. https://doi.org/10.1145/2959100.2959190

[28] Anna L Cox, Sandy J J Gould, Marta E Cecchinato, Ioanna Iacovides, and Ian Renfree. 2016. Design Frictions for Mindful Interactions: The Case for Microboundaries. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI EA '16)*. ACM, New York, NY, USA, 1389–1397. https://doi.org/10.1145/2851581.2892410

[29] David Coyle, James Moore, Per Ola Kristensson, Paul Fletcher, and Alan Blackwell. 2012. I did that! Measuring users' experience of agency in their own actions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 2025–2034. https://doi.org/10.1145/2207676.2208350

[30] Katie Davis, Anja Dinhopl, and Alexis Hiniker. 2019. " Everything's the Phone": Understanding the Phone's Supercharged Role in Parent-Teen Relationships. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. dl.acm.org, 227. https://dl.acm.org/citation.cfm?id=3300457

[31] Andy Dearden and Janet Finlay. 2006. Pattern Languages in HCI: A Critical Review. *Human–Computer Interaction* 21, 1 (March 2006), 49–102. https://doi.org/10.1207/s15327051hci2101_3

[32] Liam Delaney and Leonhard K Lades. 2017. Present bias and everyday self-control failures: a day reconstruction study. *Journal of behavioral decision making* 30, 5 (2017), 1157–1167. https://onlinelibrary.wiley.com/doi/abs/10.1002/bdm.2031?casa_token=8axLfIY-YIEAAAAA:Z65pChy95G1evF2v600EYc6oqnxIGFaC5DQJteKnCuK5AQ4Nqkj_YnMbnrJB9KxqhtTJHN0NUtLOOsI

[33] Digital Wellness Warriors. 2018. Apple: let developers help iPhone users with mental wellbeing. https://www.change.org/p/apple-allow-digital-wellness-developers-to-help-ios-users Accessed: 2020-8-27.

[34] Colin Dixon. 2019. Why shutter YouTube Leanback when there are many potential users? https://nscreenmedia.com/why-shutter-youtube-leanback-browser-experience-now/ Accessed: 2020-9-7.

[35] Angela L Duckworth, Rachel E White, Alyssa J Matteucci, Annie Shearer, and James J Gross. 2016. A Stitch in Time: Strategic Self-Control in High School and College Students. *Journal of educational psychology* 108, 3 (April 2016), 329–341. https://doi.org/10.1037/edu0000062

[36] Michael D Ekstrand and Martijn C Willemsen. 2016. Behaviorism is Not Enough: Better Recommendations Through Listening to Users. In *Proceedings of the 10th ACM Conference on Recommender Systems* (Boston, Massachusetts, USA) *(RecSys '16)*. ACM, New York, NY, USA, 221–224. https://doi.org/10.1145/2959100.2959179

[37] Robert Felner, Angela Adan, Richard Price, E L Cowen, R P Lorion, and J Ramos-McKay. 1988. 14 Ounces of prevention: A casebook for practitioners. (1988).

[38] B J Fogg. 2009. Creating persuasive technologies: an eight-step design process. In *Proceedings of the 4th International Conference on Persuasive Technology* (Claremont, California, USA) *(Persuasive '09, Article 44)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/1541948.1542005

[39] Donelson R Forsyth. 2008. Self-serving bias. (2008).

[40] Peter M Gollwitzer and Paschal Sheeran. 2006. Implementation Intentions and Goal Achievement: A Meta-analysis of Effects and Processes. In *Advances in Experimental Social Psychology*. Vol. 38. Academic Press, 69–119. https://doi.org/10.1016/S0065-2601(06)38002-1

[41] Google. [n.d.]. YouTube Leanback offers effortless viewing. https://youtube.googleblog.com/2010/07/youtube-leanback-offers-effortless.html Accessed: 2020-9-12.

[42] Colin M Gray, Yubo Kou, Bryan Battles, Joseph Hoggatt, and Austin L Toombs. 2018. The Dark (Patterns) Side of UX Design. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. ACM, New York, NY, USA, 534:1–534:14. https://doi.org/10.1145/3173574.3174108

[43] Zac Hall. 2019. Apple makes privacy extremely relatable in fun new iPhone ad - 9to5Mac. https://9to5mac.com/2019/03/14/iphone-privacy-ad/ Accessed: 2020-9-13.

[44] Jaron Harambam, Dimitrios Bountouridis, Mykola Makhortykh, and Joris van Hoboken. 2019. Designing for the Better by Taking Users into Account: A Qualitative Evaluation of User Control Mechanisms in (News) Recommender Systems. In *Proceedings of the 13th ACM Conference on Recommender Systems* (Copenhagen, Denmark) *(RecSys '19)*. ACM, New York, NY, USA, 69–77. https://doi.org/10.1145/3298689.3347014

[45] Ellie Harmon and Melissa Mazmanian. 2013. Stories of the Smartphone in Everyday Discourse: Conflict, Tension & Instability. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. ACM, New York, NY, USA, 1051–1060. https://doi.org/10.1145/2470654.2466134

[46] Joshua Hill, Kelly Widdicks, and Mike Hazas. 2020. Mapping the Scope of Software Interventions for Moderate Internet Use on Mobile Devices. In *Proceedings of the 7th International Conference on ICT for Sustainability* (Bristol, United Kingdom) *(ICT4S2020)*. Association for Computing Machinery, New York, NY, USA, 204–212. https://doi.org/10.1145/3401335.3401361

[47] Alexis Hiniker, Sharon S Heung, Sungsoo (ray) Hong, and Julie A Kientz. 2018. Coco's Videos: An Empirical Investigation of Video-Player Design Features and Children's Media Use. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18, Paper 254)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173828

[48] Alexis Hiniker, Sungsoo (ray) Hong, Tadayoshi Kohno, and Julie A Kientz. 2016. MyTime: Designing and Evaluating an Intervention for Smartphone Non-Use. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 4746–4757. https://doi.org/10.1145/2858036.2858403

[49] Alexis Hiniker, Bongshin Lee, Kiley Sobel, and Eun Kyoung Choe. 2017. Plan & Play: Supporting Intentional Media Use in Early Childhood. In *Proceedings of the 2017 Conference on Interaction Design and Children* (Stanford, California, USA) *(IDC '17)*. ACM, New York, NY, USA, 85–95. https://doi.org/10.1145/3078072.3079752

[50] Alexis Hiniker, Shwetak N Patel, Tadayoshi Kohno, and Julie A Kientz. 2016. Why would you do that? predicting the uses and gratifications behind smartphone-usage behaviors. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. dl.acm.org, 634–645. http://dl.acm.org/citation.cfm?id=2971762

[51] Wilhelm Hofmann, Roy F Baumeister, Georg Förster, and Kathleen D Vohs. 2012. Everyday temptations: an experience sampling study of desire, conflict, and self-control. *Journal of personality and social psychology* 102, 6 (June 2012), 1318–1335. https://doi.org/10.1037/a0026545

[52] Riitta Jääskeläinen. 2010. Think-aloud protocol. *Handbook of translation studies* 1 (2010), 371–374. https://books.google.com/books?hl=en&lr=&id=sBVGAYCh_9AC&oi=fnd&pg=PA371&dq=Riitta+J%C3%A4%C3%A4skel%C3%A4inen+2010+Think-aloud+protocol.&ots=Qn8NYacDXD&sig=9gAnKpZ6vdVbGHapKTfo74b0MiE

[53] Se-Hoon Jeong, Hyoungjee Kim, Jung-Yoon Yum, and Yoori Hwang. 2016. What type of content are smartphone users addicted to?: SNS vs. games. *Computers in human behavior* 54 (Jan. 2016), 10–17. https://doi.org/10.1016/j.chb.2015.07.035

[54] Anya Kamenetz. 2018. *The art of screen time: How your family can balance digital media and real life.* Hachette UK.

[55] Jaejeung Kim, Hayoung Jung, Minsam Ko, and Uichin Lee. 2019. GoalKeeper: Exploring Interaction Lockout Mechanisms for Regulating Smartphone Use. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 1 (March 2019), 29. https://doi.org/10.1145/3314403

[56] Jaejeung Kim, Joonyoung Park, Hyunsoo Lee, Minsam Ko, and Uichin Lee. 2019. LocknType: Lockout Task Intervention for Discouraging Smartphone App Use". In *ACM CHI*. https://doi.org/10.1145/3290605.3300927

[57] Young-Ho Kim, Jae Ho Jeon, Eun Kyoung Choe, Bongshin Lee, Kwonhyun Kim, and Jinwook Seo. 2016. TimeAware: Leveraging Framing Effects to Enhance Personal Productivity. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. ACM, New York, NY, USA, 272–283. https://doi.org/10.1145/2858036.2858428

[58] Minsam Ko, Subin Yang, Joonwon Lee, Christian Heizmann, Jinyoung Jeong, Uichin Lee, Daehee Shin, Koji Yatani, Junehwa Song, and Kyong-Mee Chung. 2015. NUGU: A Group-based Intervention App for Improving Self-Regulation of Limiting Smartphone Use. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Vancouver, BC, Canada) *(CSCW '15)*. ACM, New York, NY, USA, 1235–1245. https://doi.org/10.1145/2675133.2675244

[59] Geza Kovacs, Drew Mylander Gregory, Zilin Ma, Zhengxuan Wu, Golrokh Emami, Jacob Ray, and Michael S Bernstein. 2019. Conservation of Procrastination: Do Productivity Interventions Save Time Or Just Redistribute It?. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. ACM, New York, NY, USA, 330:1–330:12. https://doi.org/10.1145/3290605.3300560

[60] J R Landis and G G Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33, 1 (March 1977), 159–174. https://www.ncbi.nlm.nih.gov/pubmed/843571

[61] Gary P Latham and Edwin A Locke. 1991. Self-regulation through goal setting. *Organizational behavior and human decision processes* 50, 2 (1991), 212–247. https://www.researchgate.net/profile/Gary_Latham2/publication/232501090_A_Theory_of_Goal_Setting_Task_Performance/links/57d0e85108ae5f03b489170d/A-Theory-of-Goal-Setting-Task-Performance.pdf

[62] Paul Lewis. 2017. 'Our minds can be hijacked': the tech insiders who fear a smartphone dystopia. *The Guardian* 6, 10 (2017), 2017.

[63] Hannah Limerick, David Coyle, and James W Moore. 2014. The experience of agency in human-computer interactions: a review. *Frontiers in human neuroscience* 8 (Aug. 2014), 643. https://doi.org/10.3389/fnhum.2014.00643

[64] Hannah Limerick, James W Moore, and David Coyle. 2015. Empirical Evidence for a Diminished Sense of Agency in Speech Interfaces. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 3967–3970. https://doi.org/10.1145/2702123.2702379

[65] Danielle Lottridge, Eli Marschner, Ellen Wang, Maria Romanovsky, and Clifford Nass. 2012. Browser Design Impacts Multitasking. *Proceedings of the Human Factors and Ergonomics Society ... Annual Meeting Human Factors and Ergonomics Society. Meeting* 56, 1 (Sept. 2012), 1957–1961. https://doi.org/10.1177/1071181312561289

[66] Kai Lukoff, Alexis Hiniker, Colin M Gray, Arunesh Mathur, and Shruthi Chivukula. 2021. What Can CHI Do About Dark Patterns? In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan). ACM, New York, NY, USA. https://doi.org/10.1145/3411763.3441360

[67] Kai Lukoff, Cissy Yu, Julie Kientz, and Alexis Hiniker. 2018. What Makes Smartphone Use Meaningful or Meaningless? *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 1 (March 2018), 22:1–22:26. https://doi.org/10.1145/3191754

[68] U Lyngs, R Binns, M Van Kleek, and others. 2018. So, Tell Me What Users Want, What They Really, Really Want! *Extended Abstracts of the* (2018). https://dl.acm.org/citation.cfm?id=3188397

[69] Ulrik Lyngs, Kai Lukoff, Petr Slovak, Reuben Binns, Adam Slack, Michael Inzlicht, Max Van Kleek, and Nigel Shadbolt. 2019. Self-Control in Cyberspace: Applying Dual Systems Theory to a Review of Digital Self-Control Tools. *CHI 2019* (May 2019). https://doi.org/10.1145/3290605.3300361

[70] Ulrik Lyngs, Kai Lukoff, Petr Slovak, William Seymour, Helena Webb, Marina Jirotka, Max Van Kleek, and Nigel Shadbolt. 2020. 'I Just Want to Hack Myself to Not Get Distracted': Evaluating Design Interventions for Self-Control on Facebook. (Jan. 2020). arXiv:2001.04180 [cs.HC] http://arxiv.org/abs/2001.04180

[71] Claudia Marino, Gianluca Gini, Alessio Vieno, and Marcantonio M Spada. 2018. A comprehensive meta-analysis on Problematic Facebook Use. *Computers in human behavior* 83 (June 2018), 262–277. https://doi.org/10.1016/j.chb.2018.02.009

[72] Veronica Marotta and Alessandro Acquisti. [n.d.]. Online Distractions, Website Blockers, and Economic Productivity: A Randomized Field Experiment. ([n.d.]).

[73] Lucas Matney. 2017. YouTube has 1.5 billion logged-in monthly users watching a ton of mobile video. *TechCrunch* (June 2017). http://techcrunch.com/2017/06/22/youtube-has-1-5-billion-logged-in-monthly-users-watching-a-ton-of-mobile-video/

[74] Tom McKay. 2019. Senators Introduce Bill to Stop 'Dark Patterns' Huge Platforms Use to Trick Users. https://gizmodo.com/senators-introduce-bill-to-stop-dark-patterns-huge-plat-1833929276 Accessed: 2020-8-27.

[75] Sean M McNee, Shyong K Lam, Catherine Guetzlaff, Joseph A Konstan, and John Riedl. 2003. Confidence displays and training in recommender systems. In *Proc. INTERACT*, Vol. 3. books.google.com, 176–183. https://books.google.com/books?hl=en&lr=&id=PTg0fVYqgCcC&oi=fnd&pg=PA176&dq=low+confidence+%22recommender+system%22&ots=ObJlxzmCAZ&sig=0uf4-fGfAJyBGbfmCDJE0zpPsGU

[76] Sean M McNee, John Riedl, and Joseph A Konstan. 2006. Being accurate is not enough: how accuracy metrics have hurt recommender systems. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems* (Montréal, Québec, Canada) *(CHI EA '06)*. Association for Computing Machinery, New York, NY, USA, 1097–1101. https://doi.org/10.1145/1125451.1125659

[77] Janet Metcalfe and Matthew Jason Greene. 2007. Metacognition of agency. *Journal of experimental psychology. General* 136, 2 (May 2007), 184–199. https://doi.org/10.1037/0096-3445.136.2.184

[78] Alberto Monge Roffarello and Luigi De Russis. 2019. The Race Towards Digital Wellbeing: Issues and Opportunities. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. ACM, New York, NY, USA, 386:1–386:14. https://doi.org/10.1145/3290605.3300616

[79] James W Moore. 2016. What Is the Sense of Agency and Why Does it Matter? *Frontiers in psychology* 7 (Aug. 2016), 1272. https://doi.org/10.3389/fpsyg.2016.01272

[80] C Moser, S Y Schoenebeck, and P Resnick. 2019. Impulse Buying: Design Practices and Consumer Needs. *of the 2019 CHI Conference on ...* (2019). https://dl.acm.org/doi/abs/10.1145/3290605.3300472

[81] Arvind Narayanan, Arunesh Mathur, Marshini Chetty, and Mihir Kshirsagar. 2020. Dark Patterns: Past, Present, and Future: The evolution of tricky user interfaces. *Queueing Systems. Theory and Applications* 18, 2 (April 2020), 67–92. https://doi.org/10.1145/3400899.3400901

[82] Jakob Nielsen. 1994. 10 Heuristics for User Interface Design: Article by Jakob Nielsen. https://www.nngroup.com/articles/ten-usability-heuristics/ Accessed: 2020-2-7.

[83] NPR. 2015. Episode 653: The Anti-Store. *NPR* (Sept. 2015). https://www.npr.org/sections/money/2015/09/25/443519599/episode-653-the-anti-store

[84] Ted O'Donoghue and Matthew Rabin. 2015. Present Bias: Lessons Learned and to Be Learned. *The American economic review* 105, 5 (2015), 273–279. https://ideas.repec.org/a/aea/aecrev/v105y2015i5p273-79.html

[85] Fabian Okeke, Michael Sobolev, Nicola Dell, and Deborah Estrin. 2018. Good vibrations: can a digital nudge reduce digital overload?. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM, 4. https://doi.org/10.1145/3229434.3229463

[86] Antti Oulasvirta, Tye Rattenbury, Lingyi Ma, and Eeva Raita. 2012. Habits make smartphone use more pervasive. *Personal and Ubiquitous Computing* 16, 1 (Jan. 2012), 105–114. https://doi.org/10.1007/s00779-011-0412-2

[87] Erica Pandey. 2017. Sean Parker: Facebook was designed to exploit human "vulnerability". https://www.axios.com/sean-parker-facebook-exploits-a-vulnerability-in-humans-2507917325.html Accessed: 2020-9-15.

[88] Joonyoung Park, Jin Yong Sim, Jaejeung Kim, Mun Yong Yi, and Uichin Lee. 2018. Interaction Restraint: Enforcing Adaptive Cognitive Tasks to Restrain Problematic User Interaction. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI EA '18)*. ACM, New York, NY, USA, LBW559:1–LBW559:6. https://doi.org/10.1145/3170427.3188613

[89] Nick Paumgarten. 2014. Up And Then Down. *The New Yorker* (July 2014). https://www.newyorker.com/magazine/2008/04/21/up-and-then-down

[90] Andrew Perrin and Monica Anderson. 2019. Share of U.S. adults using social media, including Facebook, is mostly unchanged since 2018. https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/ Accessed: 2020-9-14.

[91] C Pinder, J Vermeulen, B R Cowan, and others. 2018. Digital Behaviour Change Interventions to Break and Form Habits. *ACM Transactions on* (2018). https://dl.acm.org/citation.cfm?id=3196830

[92] Paul Resnick and Hal R Varian. 1997. Recommender systems. *Commun. ACM* 40, 3 (1997), 56–58. http://dl.acm.org/citation.cfm?id=245121

[93] Kevin Roose. 2019. The making of a YouTube radical. *The New York times* (2019). https://static01.nyt.com/images/2019/06/09/nytfrontpage/scan.pdf

[94] Alan M Rubin. 1984. Ritualized and Instrumental Television Viewing. *The Journal of communication* 34, 3 (Sept. 1984), 67–77. https://doi.org/10.1111/j.1460-2466.1984.tb02174.x

[95] Richard M Ryan and Edward L Deci. 2006. Self-regulation and the problem of human autonomy: Does psychology need choice, self-determination, and will? *Journal of personality* 74, 6 (2006), 1557–1586. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-6494.2006.00420.x?casa_token=MZuzC4Br_U4AAAAA:ErU7WbByAWUFcoh2N_5TIRqe7jhVXeoV8Z0-pWB8gbb-ZZ3l8xz_qrdAtePASmiTFBWb2COf6sX4BQ

[96] Thomas C Schelling. 1984. Self-Command in Practice, in Policy, and in a Theory of Rational Choice. *The American economic review* 74, 2 (1984), 1–11. http://www.jstor.org/stable/1816322

[97] Markus Schlosser. 2019. *Agency* (winter 2019 ed.). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2019/entries/agency/

[98] Michael Schrage. 1996. Cultures of prototyping. *Bringing design to software* 4, 1 (1996), 1–11. https://pdfs.semanticscholar.org/c399/6e2738e52ea22c83ef662ab118f68b82eba2.pdf

[99] N D Schüll. 2012. *Addiction by Design: Machine Gambling in Las Vegas*. Princeton University Press. https://books.google.com/books?id=_Vsk6EXc1_4C

[100] Nick Seaver. 2018. Captivating algorithms: Recommender systems as traps. *Journal of Material Culture* (Dec. 2018), 1359183518820366. https://doi.org/10.1177/1359183518820366

[101] Ben Shneiderman. 1992. *Designing the user interface (2nd ed.): strategies for effective human-computer interaction*. Addison-Wesley Longman Publishing Co., Inc., USA. https://dl.acm.org/citation.cfm?id=129385

[102] Ben Shneiderman and Catherine Plaisant. 2004. *Designing the User Interface: Strategies for Effective Human-Computer Interaction (4th Edition)*. Pearson Addison Wesley.

[103] Jonathan Sillito. [n.d.]. Saturate App: Simple Collaborative Analysis. http://www.saturateapp.com/ Accessed: 2020-2-NA.

[104] L Silver, A Smith, C Johnson, K Taylor, J Jiang, A Monica, and L Rainie. 2019. Use of smartphones and social media is common across most emerging economies. https://www.pewresearch.org/internet/2019/03/07/use-of-smartphones-and-social-media-is-common-across-most-emerging-economies/ Accessed: 2019-2-NA.

[105] Aaron Smith, Skye Toor, and Patrick Van Kessel. 2018. Many Turn to YouTube for Children's Content, News, How-To Lessons. https://www.pewresearch.org/internet/2018/11/07/many-turn-to-youtube-for-childrens-content-news-how-to-lessons/ Accessed: 2020-3-3.

[106] Falko F Sniehotta, Urte Scholz, and Ralf Schwarzer. 2005. Bridging the intention–behaviour gap: Planning, self-efficacy, and action control in the adoption and maintenance of physical exercise. *Psychology & Health* 20, 2 (April 2005), 143–160. https://doi.org/10.1080/08870440512331317670

[107] Joan E Solsman. 2018. Ever get caught in an unexpected hourlong YouTube binge? Thank YouTube AI for that. https://www.cnet.com/news/youtube-ces-2018-neal-mohan/ Accessed: 2020-5-1.

[108] Todd Spangler. [n.d.]. YouTube Tops 20 Million Paying Subscribers, YouTube TV Has Over 2 Million Customers. https://variety.com/2020/digital/news/youtube-tops-20-million-paying-subscribers-youtube-tv-has-over-2-million-customers-1203491228/ Accessed: 2020-8-26.

[109] Maggie Stanphill. 2019. Optimizing for Engagement: Understanding the Use of Persuasive Technology on Internet Platforms. U.S. Senate Committee Hearing.

[110] Nick Statt. 2016. Flowstate is a writing app that will delete everything if you stop typing. https://www.theverge.com/2016/1/28/10853534/flowstate-writing-app-mac-ios-delete-everything Accessed: 2020-8-10.

[111] S Shyam Sundar and Sampada S Marathe. 2010. Personalization versus Customization: the Importance of Agency, Privacy, and Power Usage. *Human communication research* 36, 3 (July 2010), 298–322. https://doi.org/10.1111/j.1468-2958.2010.01377.x

[112] Matthis Synofzik, Gottfried Vosgerau, and Albert Newen. 2008. Beyond the comparator model: a multifactorial two-step account of agency. *Consciousness and cognition* 17, 1 (March 2008), 219–239. https://doi.org/10.1016/j.concog.2007.03.010

[113] Adam Tapal, Ela Oren, Reuven Dar, and Baruch Eitam. 2017. The Sense of Agency Scale: A Measure of Consciously Perceived Control over One's Mind, Body, and the Immediate Environment. *Frontiers in psychology* 8 (Sept. 2017), 1552. https://doi.org/10.3389/fpsyg.2017.01552

[114] Jonathan A Tran, Katherine S Yang, Katie Davis, and Alexis Hiniker. 2019. Modeling the Engagement-Disengagement Cycle of Compulsive Phone Use. In *CHI '19* (Glasgow). https://doi.org/10.1145/3290605.3300542

[115] United States Census Bureau. [n.d.]. QuickFacts: United States. https://www.census.gov/quickfacts/fact/table/US/PST045219

[116] Patrick Van Kessel, Skye Toor, and Aaron Smith. 2019. A Week in the Life of Popular YouTube Channels. https://www.pewresearch.org/internet/2019/07/25/a-week-in-the-life-of-popular-youtube-channels/ Accessed: 2020-4-1.

[117] James Williams. 2018. *Stand Out of Our Light: Freedom and Resistance in the Attention Economy*. Cambridge University Press. https://play.google.com/store/books/details?id=88FWDwAAQBAJ

[118] Eva Yiwei Wu, Emily Pedersen, and Niloufar Salehi. 2019. Agent, Gatekeeper, Drug Dealer: How Content Creators Craft Algorithmic Personas. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW (Nov. 2019), 219:1–219:27. https://doi.org/10.1145/3359321

[119] YouTube. [n.d.]. YouTube for Press. https://www.youtube.com/about/press/ Accessed: 2020-8-14.

[120] José P Zagal, Staffan Björk, and Chris Lewis. 2013. Dark Patterns in the Design of Games. In *Foundations of Digital Games 2013*. diva-portal.org. http://www.diva-portal.org/smash/record.jsf?pid=diva2:1043332

[121] John Zimmerman and Jodi Forlizzi. 2014. Research Through Design in HCI. In *Ways of Knowing in HCI*, Judith S Olson and Wendy A Kellogg (Eds.). Springer New York, New York, NY, 167–189. https://doi.org/10.1007/978-1-4939-0378-8_8

Int. J. Human-Computer Studies 190 (2024) 103303

Contents lists available at ScienceDirect

International Journal of Human - Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

Preventing users from going down rabbit holes of extreme video content: A study of the role played by different modes of autoplay

Cheng Chen[a,*], Jingshi Kang[b], Pejman Sajjadi[c], S. Shyam

[a] Elon University, Elon, NC, USA
[b] Fudan University, Shanghai, PR China
[c] Meta, California, USA
[d] Pennsylvania State University, University Park, PA, USA

Autoplay

ARTICLE INFO

Keywords:
Interpassivity
Autoplay
Control heuristic
Content extremity
Inattentiveness
Rabbit hole perception
Negative expectancy violation

ABSTRACT

The autoplay feature of video pla ... tching
extreme content. However, autopl ... gle the
feature off if they want. While the a ... king it
"interpassive," which lies betwee ... o. We
empirically compare these three mo ... ants to
either extreme or non-extreme con ... pletely
passive autoplay. Results show the ... control
heuristic compared to passive autop ... th the
invoked control heuristic and inatt ... ve im-
plications for socially responsible d

1. Introduction

Content extremity is a long-standing issue on YouTube. Its recommendation algorithms expose users to progressively more extreme content (McCrosky and Geurkink, 2021; Ribeiro et al., 2020; Tang et al., 2021; Tufekci, 2018). For example, an algorithmic audit by Tufekci (2018) revealed that YouTube videos about Donald Trump rallies led to white supremacist rants and Holocaust denials. Videos about Hillary Clinton and Bernie Sanders moved to those about secret government agencies. Even ideologically neutral videos gradually take an extreme turn, changing from jogging to running, and from marathons to ultramarathons. The YouTube Regrets report also documents user-reported examples where the recommendation algorithms led users to see socially undesirable and even harmful content containing misinformation and violent or sexually explicit content (McCrosky and Geurkink, 2021).

Researchers use the phrase "down the rabbit hole" to describe users' encounters with extreme content on YouTube (Rosen et al., 2022; Kaiser, 2019; Ledwich and Zaitsev, 2019; Tang et al., 2021), and tend to blame the feature of autoplay for this phenomenon (Giuffredi and De Rosnis, 2022; Tang et al., 2021). If this is true, then altering autoplay should help avoid potential harms caused by exposure to extreme content (Livingstone et al., 2014; Tang et al., 2021). We investigate this possibility through the lens of interpassivity (Pfaller, 1996; Zizek, 1997) and examine how different modes of autoplay affect users' rabbit hole perception under conditions of both extreme and non-extreme video recommendations.

Interpassivity affords the action possibility of delegating a task to an agent, enabling a paradoxical experience that combines both passivity (due to automation) and interactivity (due to user control). Does the automation aspect trigger the "machine heuristic," a perception that machines are more objective than humans (Sundar and Kim, 2019), thus less likely to drive users down a rabbit hole, given that new content is automatically selected by recommendation algorithms? Does automation leave users inattentive to what they are watching considering that they did not actively choose the video content? Does the aspect of interactivity provide a heightened sense of user control, considering that it allows users to toggle the feature on and off? Answering these questions could help unpack the psychological effect of autoplay in the context of online video viewing.

Studies have pointed out that being stuck in a rabbit hole is a complex experience, which could also be attributed to the consecutiveness of one's prior media consumption experience (Woolley and Sharif, 2022).

* Correspondence author at: Communication Design Department, School of Communications, Elon University, Elon, NC, 27244
E-mail address: cchen8@elon.edu (C. Chen).

https://doi.org/10.1016/j.ijhcs.2024.103303
Received 8 November 2023; Received in revised form 23 March 2024; Accepted 30 May 2024
Available online 5 June 2024
1071-5819/Published by Elsevier Ltd.

Int. J. Human–Computer Studies 190 (2024) 103303

Contents lists available at ScienceDirect

International Journal of Human - Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

Preventing users from going down rabbit holes of extreme video content: A study of the role played by different modes of autoplay

Cheng Chen[a,*], Jingshi Kang[b], Pejman Sajjadi[c], S. Shyam Sundar[d]

[a] Elon University, Elon, NC, USA
[b] Fudan University, Shanghai, PR China
[c] Meta, California, USA
[d] Pennsylvania State University, University Park, PA, USA

ARTICLE INFO

Keywords:
Interpassivity
Autoplay
Control heuristic
Content extremity
Inattentiveness
Rabbit hole perception
Negative expectancy violation

ABSTRACT

The autoplay feature of video platforms is often blamed for users going down rabbit holes of binge-watching extreme content. However, autoplay is not necessarily a passive experience, because users can toggle the feature off if they want. While the automation aspect is passive, the toggle option signals interactivity, making it "interpassive," which lies between completely passive autoplay and manual initiation of each video. We empirically compare these three modes of video viewing in a user study ($N = 394$), which exposed participants to either extreme or non-extreme content under conditions of manual play, interpassive autoplay, or completely passive autoplay. Results show that interpassive autoplay is favored over the other two. It triggers the control heuristic compared to passive autoplay, but leads to higher inattentiveness compared to manual play. Both the invoked control heuristic and inattentiveness result in higher rabbit hole perception. These findings have implications for socially responsible design of the autoplay feature.

1. Introduction

Content extremity is a long-standing issue on YouTube. Its recommendation algorithms expose users to progressively more extreme content (McCrosky and Geurkink, 2021; Ribeiro et al., 2020; Tang et al., 2021; Tufekci, 2018). For example, an algorithmic audit by Tufekci (2018) revealed that YouTube videos about Donald Trump rallies led to white supremacist rants and Holocaust denials. Videos about Hillary Clinton and Bernie Sanders moved to those about secret government agencies. Even ideologically neutral videos gradually take an extreme turn, changing from jogging to running, and from marathons to ultramarathons. The YouTube Regrets report also documents user-reported examples where the recommendation algorithms led users to see socially undesirable and even harmful content containing misinformation and violent or sexually explicit content (McCrosky and Geurkink, 2021).

Researchers use the phrase "down the rabbit hole" to describe users' encounters with extreme content on YouTube (Brown et al., 2022; Kaiser, 2019; Ledwich and Zaitsev, 2019; Tang et al., 2021), and tend to blame the feature of autoplay for this phenomenon (Buffarello and De Russis, 2022; Tang et al., 2021). If this is true, then altering autoplay should help avoid potential harms caused by exposure to extreme content (Livingstone et al., 2014; Tang et al., 2021). We investigate this possibility through the lens of interpassivity (Pfaller, 1996; Žižek, 1997) and examine how different modes of autoplay affect users' rabbit hole perception under conditions of both extreme and non-extreme video recommendations.

Interpassivity affords the action possibility of delegating a task to an agent, enabling a paradoxical experience that combines both passivity (due to automation) and interactivity (due to user control). Does the automation aspect trigger the "machine heuristic," a perception that machines are more objective than humans (Sundar and Kim, 2019), thus less likely to drive users down a rabbit hole, given that new content is automatically selected by recommendation algorithms? Does automation leave users inattentive to what they are watching considering that they did not actively choose the video content? Does the aspect of interactivity provide a heightened sense of user control, considering that it allows users to toggle the feature on and off? Answering these questions could help unpack the psychological effect of autoplay in the context of online video viewing.

Studies have pointed out that being stuck in a rabbit hole is a complex experience, which could also be attributed to the consecutiveness of one's prior media consumption experience (Woolley and Sharif, 2022).

* Correspondence author at: Communication Design Department, School of Communications, Elon University, Elon, NC, 27244
E-mail address: cchen8@elon.edu (C. Chen).

https://doi.org/10.1016/j.ijhcs.2024.103303
Received 8 November 2023; Received in revised form 23 March 2024; Accepted 30 May 2024
Available online 5 June 2024
1071-5819/Published by Elsevier Ltd.

Users are more likely to choose similar media content over dissimilar media content or completing a non-media task after consecutively consuming multiple pieces of similar content (Woolley and Sharif, 2022). Thus, we are motivated to explore how users' prior online video viewing influences their perceptions of extreme video content under autoplay.

To answer these questions, we designed a user study in which we randomly assigned participants to use a self-developed online video platform in three different autoplay modes: interpassive autoplay, manual play, and completely passive autoplay. In each autoplay mode, participants watched videos that were either progressively more extreme or similar to each other. After this, we measured their rabbit hole perception and user experience related to the use of autoplay feature. We found that interpassive autoplay increases users' perception of going down the rabbit hole under certain conditions and further revealed two psychological mechanisms that drive the effect of autoplay on this perception.

This study contributes to the literature in the following ways: First, drawing on the perspective of interpassivity, it reveals the nature of autoplay by demonstrating how automation and interactivity, as two independent affordances, differently affect users' perceptions and experience with the autoplay feature in online video platforms. Second, the study reveals the control heuristic and inattentiveness as two psychological mechanisms underlying the effect of autoplay on rabbit hole perception. Ascertaining the role of these mediators deepens our understanding of the psychological effect of autoplay in forming users' rabbit hole perception. Lastly, this study reveals an individual difference—the amount of prior online video viewing—as influencing the effect of autoplay on rabbit hole perception. Such nuanced findings contribute to the conditional media effects paradigm by revealing the conditions under which the effect occurs. In addition to the theoretical implications, findings of the study inform interface designs for minimizing the potential risks of engaging with algorithmic personalization that may expose users to extreme content and improving algorithmic literacy among users to promote mindful and meaningful interactions on online video platforms.

2. Related work

In this section, we first introduce three modes of autoplay based on the concept of interpassivity and discuss their potential effects on user experience (UX) and user interface (UI) satisfaction. Following this, we explore the effects of autoplay on users' rabbit hole perception. By focusing on two distinct affordances, automation and interactivity, we discuss the psychological mechanisms that may influence rabbit hole perception under different autoplay modes. After presenting each key study concept, we propose research questions and hypotheses.

2.1. *Autoplay and affordance of interpassivity*

Autoplay has been a feature on YouTube since 2015. It automatically queues and plays a sequence of videos based on the preferences of users themselves and also those of collective others, thereby streamlining content consumption. It sustains user engagement by reducing the need for manual video selection but can also lead to unintended content exposure due to its algorithmic nature. The autoplay feature is enabled on YouTube by default, but users can disable the feature any time using the toggle option available on the video-playing page. The two dimensions, automation and interactivity, collectively shape user experience with autoplay. Early studies have conceptualized this unique affordance as interpassivity (Pfaller, 1996; Žižek, 1997), which has become a new source of gratification (Chen et al., 2023).

As an automated feature, autoplay is enjoyed by users as it affords convenience (by freeing users from doing things manually), user control (it allows users to have the final say), and user profiling (it remembers users' habits and preferences) (Chen et al., 2023). But it has also been criticized as a dark-pattern design feature because it undermines users' sense of agency (Lukoff et al., 2021) and promotes excessive technology use and problematic behavior (Ruffarello and De Russis, 2022; Schaffner et al., 2023; Stöcker and Preuss, 2020).

Given that these studies take an object-centered approach by treating autoplay as a holistic object, we do not know which aspects of autoplay contribute to these unintended consequences. Scholars have argued that the object-centered approach can inhibit knowledge generation (Nass and Mason, 1990) as it may be difficult to generalize findings about YouTube's autoplay to other platforms that use the same feature. To better understand the effect of autoplay and make it generalizable to other technologies that adopt the same mechanisms, we take a variable-centered approach by specifying the affordances underlying the autoplay feature, i.e., automation and interactivity.

On the one hand, autoplay affords automation. Users can delegate the video selection task to the machine and enjoy having the videos automatically played for them. The task delegation may provide a sense of relaxation and convenience. On the other hand, autoplay affords interactivity by allowing users to turn the autoplay feature off. The action possibility provided by interactive features, such as the toggle option, may result in a heightened sense of user control. Depending on users' engagement with these two affordances, there are three modes of autoplay for consideration in the current study context: interpassive autoplay (automation + interactivity), manual play (interactivity only), and completely passive autoplay (automation only). By comparing interpassive autoplay with manual play, we can observe the psychological effects of automation on user experience and perceptions of the recommended content. Furthermore, the distinction between interpassive autoplay and completely passive autoplay allows us to evaluate the effect of interactivity on perceptual outcomes. As such, the first goal of our study is to explore the effects of different autoplay modes on user experience (UX) and user interface (UI) satisfaction through the lens of interpassivity.

2.2. *UX and UI satisfaction with autoplay*

We define UX as "the person's experience at the moment experienced" (Whiteside and Wixon, 1987). It not only concerns usability issues, such as productivity and learnability of the system, but also focuses on the amount of fun and enjoyment users perceive in their interaction with it. Overall, user experience is a holistic perception obtained from interaction with the system. It could be hedonic, affective, or experiential aspects of technology use (Hassenzahl and Tractinsky, 2006). UI is one aspect that influences user experience. Compared to UX, UI focuses more on the elements presented on the interface, such as icons, buttons, and content, and how the presence of certain or all elements influences users' reactions to the system (Chin et al., 1988).

As an aspect of UI, the autoplay feature has fundamentally changed user experience with online video platforms such as YouTube. Early studies on autoplay have pointed out several reasons why users like (Chen et al., 2023) and dislike this feature (Lukoff et al., 2021; Ruffarello et al., 2023; Ruffarello and De Russis, 2022). The reasons for dislike include usurping users' sense of agency (Lukoff et al., 2021, 2023), the absence of easy-to-find buttons to turn off the feature (Ruffarello et al., 2023), promoting prolonged use (Ruffarello and De Russis, 2022; Schaffner et al., 2023), and leading users down the rabbit hole through personalization algorithms (McCrosky and Geurkink, 2021). However, these studies are mostly based on interviews, focus groups, and survey data, thus making it difficult to infer causation. In other words, we do not know whether autoplay negatively affects UX and UI satisfaction, and if so, which aspects of the autoplay experience contribute to these negative outcomes. Unpacking the affordances provided by autoplay through the lens of interpassivity and employing an experimental method enable us to more effectively observe the psychological effect of autoplay compared to pre-existing play modes, such as manual play.

Furthermore, some people may use autoplay in a completely passive

way by not intervening at all in the display of the videos. Previous research has pointed out that users do not seem to enjoy being passive viewers, probably due to the lack of user control over video recommendations (Raffaello and De Rosis, 2022). However, it remains unknown how the UX and UI satisfaction differ across interpassive autoplay and completely passive autoplay given the absence of an apple-to-apple comparison. The lack of empirical study comparing different modes of autoplay in UX and UI satisfaction motivates us to ask the following research question:

RQ1. How does interpassive autoplay influence UX and UI satisfaction compared to manual play and completely passive autoplay?

2.3. *Autoplay, content extremity, rabbit hole perception*

Beyond UX and UI satisfaction, a major concern about autoplay is its potential to shape users' reception and perceptions of extreme online content. Given that autoplay has long been referred to as a "radicalizer" for its tendency to progressively recommend and supply more extreme content (Ribeiro et al., 2020; Tang et al., 2021; Tufekci, 2018), our second study goal is to examine whether users perceive themselves as falling into a rabbit hole when exposed to increasingly extreme videos under autoplay.

Conceptually, we define content extremity as recommendations on daily life activities escalating towards extremes. One example is videos on running/jogging evolving into marathon and ultramarathon in Tufekci's (2018) algorithmic audit. We focus on daily life activities—avoiding socially undesirable and harmful content featuring radical political ideologies or misinformation—to more accurately reflect the typical user experience on online video platforms. The top three most watched video genres on YouTube are music videos, tutorials, and "top" list about makeup or dinner recipes (10 Most Watched Categories of YouTube Videos, 2022). By focusing on daily-life video content that progressively become more extreme, our study also aims to enhance external validity, considering that any video topic or theme can potentially veer towards extremism, thereby illustrating the rabbit hole phenomenon.

Realizing that recommended contents become progressively more extreme is crucial for users' digital well-being, as it paves the way for mindful media usage. Moreover, when algorithms recommend content like false news or hate speech, recognizing the potential for falling into a rabbit hole can act as a safeguard. It may prevent users from engaging with, or at the very least being inoculated against, misleading, inflammatory, and even harmful content (Kaiser, 2019; Ledwich and Zaitsev, 2019; Tang et al., 2021).

We term this positive cognitive outcome as *rabbit hole perception*, which is defined as a cognitive state in which users are alert to the nature of the recommendation algorithm, particularly its tendency to escalate towards more extreme manifestations of the initially chosen topic or theme, often to the point of differing significantly from users' initial viewing goal. An absence of this perception means users are insufficiently aware of this discrepancy between their goals and the videos being fed to them. They are lulled into overlooking the negative violation of their expectations. That is why a conscious perception of being led down the rabbit hole is important for stemming the negative effects of mindless watching. It can serve as a wake-up call, alerting users to scrutinize the nature of the recommendations. This means an awareness of negative expectancy violation is a logical antecedent of rabbit hole perception.

The question is whether this perception can be better achieved with interpassive autoplay compared to manual play and completely passive autoplay when the video platform progresses toward extreme content. We examine the affordance of interpassivity and discuss how each aspect of the affordance, i.e., automation and interactivity, may influence user experience and perceptions of extreme content, respectively, by invoking different cognitive heuristics, as discussed in the sections below.

2.3.1. *Psychology of automation: inattentiveness and machine heuristic*

Automation allows users to enjoy videos without the tedium of frequently interacting with the interface. The lower task demand under the completely passive autoplay mode may cause mental underload, a status where one's attentional resources are not sufficiently activated to process the information at hand (Young and Stanton, 2002). Researchers have pointed out that mental underload is not an optimal mental status because it can negatively influence user performance under emergencies (Young and Stanton, 2002). It suggests that disengagement in the face of video content, or inattentiveness, may reduce vigilance to a point where users do not realize that their expectations are being violated, thus undermining their perception of going down a rabbit hole. Based on this rationale, we predict that users may be less likely to perceive they are going down the rabbit hole when they use the autoplay feature. This is more likely to be true in the interpassive autoplay condition compared to the manual play condition because the lower task demand of the former results in mental underload. Therefore, we hypothesize:

H1. Compared to manual play, interpassive autoplay will trigger higher inattentiveness, which will be associated with less negative expectancy violation, leading to lower rabbit hole perception when the recommended video content becomes extreme.

Another possible psychological outcome of automation is the triggering of "machine heuristic," which is a user perception that machines are more objective and neutral than humans (Sundar and Kim, 2019). This perception is often formed when the source of interaction is a machine, and users tend to apply stereotypical thinking about machines to understand the interaction at hand (Sundar, 2008; Sundar and Kim, 2019). We argue that the presence of the autoplay feature may activate the machine heuristic because the source of recommendation is an algorithm rather than a human (Markmann and Grimme, 2021; Stöcker and Preuss, 2020). Users are likely to overtrust the system's ability to produce quality recommendations and therefore let their antenna down in terms of critically evaluating the relevance of the videos to their initial interest. As a result, they may not be aware of the negative violation of their expectations, resulting in lower rabbit hole perception. Thus, we propose the following hypothesis:

H2. Compared to manual play, interpassive autoplay will activate the machine heuristic, which will be associated with less negative expectancy violation, leading to lower rabbit hole perception when the recommended video content becomes extreme.

2.3.2. *Psychology of interactivity: control heuristic*

Independent of automation, autoplay also affords interactivity, which allows users to toggle the feature on and off. The sheer presence of the toggle button and other interaction possibilities may invoke the control heuristic, promoting the belief among users that if they are in charge of the video being played, the recommended video must be good (Sundar, 2008). This belief implies an ego defensiveness effect (Miller, 1976). That is, once users feel in charge of the functionality of the platform, they tend to engage in a self-enhancing and defensive processing of the recommended video content. The positive perception toward the recommended content may reduce negative expectancy violation, resulting in less rabbit hole perception. Based on this rationale, we propose the following hypothesis:

H3. Compared to completely passive autoplay, interpassive autoplay will trigger the control heuristic, which will be associated with less negative expectancy violation, leading to lower rabbit hole perception when the recommended video content becomes extreme.

2.4. *Individual difference: amount of prior online video viewing*

Individual differences can affect users' perception of extreme

content. We propose that heavy online video viewers may be less aware that they are in a rabbit hole when using the interpassive autoplay feature. There are two distinct but related concepts that can explain this prediction. One pertains to the notion of desensitization, which states that individuals tend to have diminished emotional responsiveness to stimuli after repeated exposure (Carnagey et al., 2007). Given that online video platforms, such as YouTube, often feature all kinds of videos, including radical and extreme content (Ribeiro et al., 2020; Tang et al., 2021), the more time users spend watching online videos, the less sensitive they are toward extreme content, thus reducing rabbit hole perception. This tendency is hypothesized to be stronger for those who are under the interpassive autoplay mode due to the lack of attention or the invoked machine heuristic, as discussed earlier.

Related to the desensitization point of view, the lack of rabbit hole perception could also be explained by cultivation theory, which posits that long-term media exposure can distort individuals' perception of reality (Gerbner, 1967, 1969). Considering that heavy viewers tend to spend longer time watching videos online, they may encounter extreme content more often than an average viewer. As a result, they may view rare and novel content recommendations as common, popular, and mainstream. Based on this rationale, heavy online video viewers may be less aware that they are in a rabbit hole when the platform features extreme video content. Again, this effect is likely to be stronger when the interpassive autoplay is on, as the automation aspect can trigger inattentiveness and machine heuristic.

Together, we propose the moderating effect of heavy online video viewing on the interaction between autoplay modes and content extremity on rabbit hole perception, as shown below.

H4. The proposed effect in H1 will be moderated by the amount of prior online video viewing, such that the effect hypothesized in H1 will be stronger for those who are heavy, compared to light, online video viewers.

H5. The proposed effect in H2 will be moderated by the amount of prior online video viewing, such that the effect hypothesized in H2 will be stronger for those who are heavy, compared to light, online video viewers.

We present the study model in Fig. 1.

3. Method

To address the research question and test the hypotheses, we conducted a 3 (autoplay modes: interpassive autoplay vs. manual play vs. completely passive autoplay) × 2 (content extremity: non-extreme vs. extreme) between-subjects online experiment. We obtained the University's Institutional Review Board approval and pre-registered the study on open science foundation (OSF) before data collection. Click here https://osf.io/s7jkr to see the details of the pre-registration.

3.1. Participants

An *a priori* power analysis indicated that at least 279 participants were required to achieve the power of 0.80, an error rate of 0.05, and a small to medium effect size (f = 0.25) in the F family tests. Given that there may be invalid and inattentive responses, we recruited 400 participants from Cloud Research to ensure sufficient power for data analysis. After excluding participants who failed either of the two attention checks, we were left with 394 participants.

Our sample consisted of slightly more females (50.3 %) than males (47.4 %). 1.5 % of participants chose the non-binary gender category, and 0.5 % preferred not to answer this question. Their ages ranged from 21 to 85 years old (M = 42.76, SD = 12.69). The median family income was $50,000 to less than $75,000 per year, and the median educational background was a bachelor's degree. Our sample was dominated by White participants (71.3 %), followed by Blacks or African Americans (17 %), Asians (7.6 %), American Indians or Alaska Natives (1.8 %), Native Hawaiians or Pacific Islanders (5 %). It is worth noting that 0.8 % preferred not to reveal their race background and 1.5 % chose other races or origins.

3.2. Stimuli

We created 12 versions of an online video platform prototype for the study. Participants were informed that a newly developed online video platform called VIDNATION provides high-quality videos tailored to users' interests. They were invited to try out the beta version, with the reminder that the platform was in the prototype phase and some features might not be fully functional. After the introduction, participants were provided a link to access the online video platform. Depending on their assigned condition, they were asked to search either "jogging" or "meal prep" to get started. They were informed that the platform would show them a total of four videos, and they would need to provide a four-digit code to proceed after returning from the platform.

As shown in Fig. 2, participants first landed on the home page of VIDNATION, where they were shown a list of 16 videos on different topics. After entering the required search query ("jogging" or "meal prep"), participants were shown the search results page, which displayed 12 video thumbnails. The thumbnails were generic and ambiguous, so users could not clearly predict the upcoming video content. Participants could click on any of the 12 videos to proceed. Regardless of which video they chose, it directed them to the same next page.

A tour was provided before video playing to introduce the functionality of each major feature on the platform, including the basic video

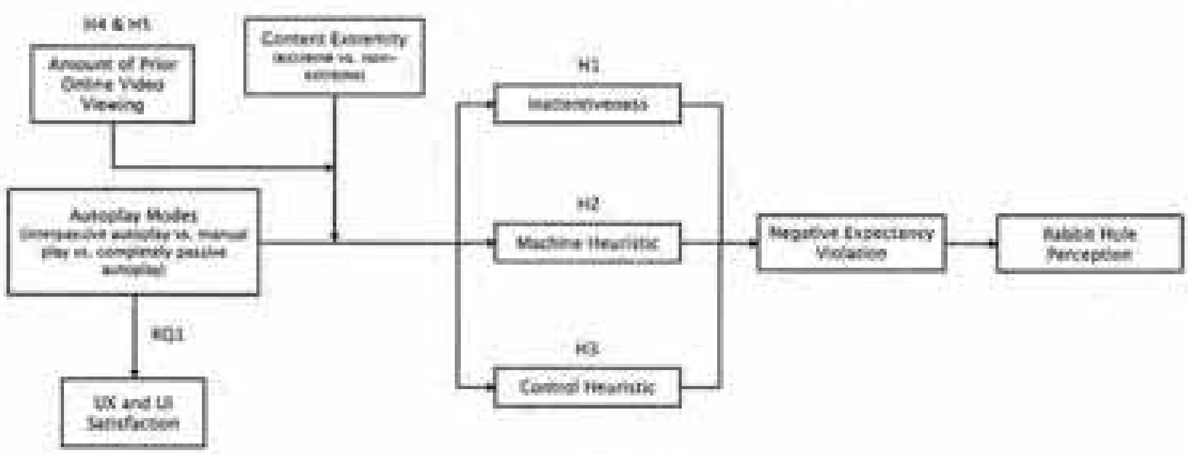


Fig. 1. Study Model.

control buttons (i.e., the play button, the progress bar, and the volume control button), the Next Video button (i.e., click to see the next video), and the autoplay toggle option (i.e., turn the autoplay feature on and off). We provided this tour because our pretests showed that users could not clearly differentiate interpassive autoplay and manual play in terms of user control. They perceived both play modes to be similar in user control and significantly higher than the user control offered in the completely passive autoplay condition. The pretest result indicated that simply providing an autoplay toggle option on the interface might be too subtle. Thus, we added the tour to help users understand the functionality of each feature, a point further discussed in the practical implication section.

Following the tour, participants were shown four videos in a predefined order so that they were exposed to either extreme or non-extreme video content. Each video lasted approximately one minute. By the end of the last video, participants were provided with a four-digit access code to proceed with the rest of the questionnaire. We describe the manipulation of autoplay modes and content extremity in the subsections below.

3.2.1. Manipulation of autoplay modes

We manipulated three modes of autoplay based on the presence or absence of three interface features, including toggling the autoplay feature on and off, controlling the Next Video button, and having the platform automatically play the next video (see Table 1). Given that autoplay itself affords interpassivity, i.e., both interactivity and automation, the interpassive autoplay condition not only allowed users to enjoy the videos automatically played one after another with a 3-second countdown between each, but also allowed users to turn the autoplay feature off through the toggle option if they desired. To increase ecological validity, we also made the "Next Video" button clickable, so that users could jump to the next video if they were not satisfied with the autoplay recommendation.

By contrast, the manual play condition only offered user control over the Next Video button but disabled the toggle option. Given that the automation aspect was not available in manual play, there was no 3-second countdown between videos. The completely passive autoplay represented a real-life situation where users just lean back and enjoy the video automatically played for them without taking any actions. Thus, we had the platform automatically play the next video. Also, we disabled the toggle option and the Next Video button in the completely passive autoplay condition.

3.2.2. Manipulation of content extremity

To manipulate content extremity, we used stimulus sampling by randomly assigning two video topics to our participants for viewing: jogging and meal prep. The purpose of stimulus sampling is to increase both the internal and external validity of the study, as content extremity could vary across topics. For the topic of jogging, the recommended videos in the extreme condition progressed from jogging to 5K running, and then to marathon, and finally to ultramarathon. In contrast, videos in the non-extremity condition were similar to each other, focusing on

Fig. 2. Prototype of the online video platform in the manual play condition.

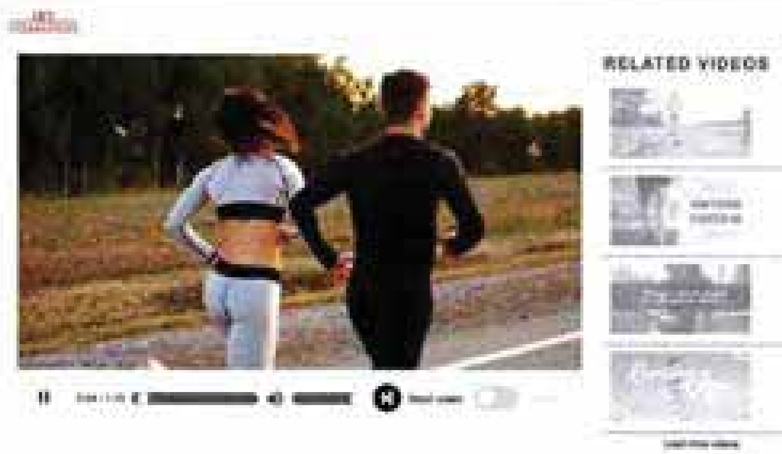


Fig. 2. (continued).

Table 1
Manipulation of autoplay modes.

Autoplay Modes	Toggling the autoplay feature on and off	Controlling the Next Video button	Having the platform automatically play the next video with a 3-second countdown
Interpassive Autoplay	Yes	Yes	Yes
Manual Play	No	Yes	No
Completely Passive autoplay	No	No	Yes

the effects and tips of jogging and running.

Regarding meal prep, we manipulated content extremity by increasing the number of people served in a meal from a single person to a family of four, and then moving from serving 1,000 people to 100,000 people. By contrast, the non-extreme condition featured cooking videos with dishes commonly seen in American society, such as mustard chicken, chicken breast, salmon, and pasta. In both the jogging and meal prep video series, both extreme and non-extreme video conditions started with the same first video.

3.3. *Procedure*

Participants went through three stages in the study: 1) filling out a pre-test questionnaire, 2) interacting with the online video platform VIDNATION, and 3) completing a post-test questionnaire. In the pre-test questionnaire, we measured their demographics and amount of prior online video viewing. Then, they were randomly assigned to interact with one of the 12 versions of VIDNATION. After this, they completed the post-test questionnaire, which included manipulation check questions and measures for all mediators and dependent variables, as described below.

3.4. *Measures*

We measured the following four mediators: inattentiveness, machine heuristic, control heuristic, and negative expectancy violation. The dependent variables were rabbit hole perception, UX, and UI satisfaction. The moderating variable was the amount of prior online video viewing. All measures were rated on a 7-point scale, unless otherwise explained.

3.4.1. *Inattentiveness*

Inattentiveness was measured by four items adapted from prior research (LaRose et al., 2003), including 1) I did not pay much attention to how VIDNATION recommended videos to me; 2) My mind wandered when VIDNATION recommended videos to me; 3) I did not really think about how VIDNATION recommended video; and 4) I was not aware of how VIDNATION recommended video. We created an index of inattentiveness by averaging the four items ($M = 2.36$, $SD = 1.39$), which

was reliable Cronbach's *alpha* = 0.89.

3.4.2. Machine heuristic

Measures for machine heuristic were modified from earlier studies (Sundar and Kim, 2019; Yang and Sundar, 2020). Items included: 1) VIDNATION is better than humans in making recommendations about videos to watch; 2) VIDNATION is more dependable than humans in selecting the best next video; 3) VIDNATION is more reliable than humans in video recommendations; and 4) VIDNATION is more competent than humans in searching for the best next video. The average of the four items formed an index of the machine heuristic (M = 3.67, SD = 1.49), which was reliable, Cronbach's *alpha* = 0.96.

3.4.3. Control heuristic

Based on the conceptualization of the control heuristic in previous studies (Sundar, 2008; Sundar et al., 2020), we developed three items to fit the context of online video watching, namely 1) VIDNATION allowed me to control the play of the videos, so the platform must be good; 2) I felt in charge of my video watching experience, so the platform must be good; and 3) VIDNATION afforded me control over video playing, so the platform must be good. The three-item measure was also highly reliable (Cronbach's *alpha* = 0.95, M = 4.09, SD = 1.86).

3.4.4. Negative expectancy violation

Modified from earlier studies (Ramirez Jr and Wang, 2008; Waddell, 2018), we used three items to evaluate the extent to which users' expectations were violated by the recommended video content. Items included: 1) I was disappointed by the videos played on VIDNATION, 2) The videos were not as good as I thought they would be, and 3) The videos did not meet my expectations. An index was created by averaging the three items, and it was quite reliable (Cronbach's *alpha* = 0.95, M = 3.26, SD = 1.64).

3.4.5. Rabbit hole perception

Given the absence of a well-established scale to measure rabbit hole perception, we developed four items to assess the extent to which users perceive that the recommended video content went to an extreme. The items were: 1) The videos shown to me were progressively more extreme; 2) The videos shown to me were more and more extreme; 3) The videos took me down into a rabbit hole; and 4) The videos shown to me were different from what I searched in the beginning. To identify the latent construct of the four items, we conducted a principal component factor analysis with Varimax rotation. The analysis yielded a single factor with an eigenvalue greater than 1, with our four items explaining 68.81 % of the variance in rabbit hole perception. Thus, we created an index by averaging the four items (Cronbach's *alpha* = 0.85, M = 2.66, SD = 1.39). Recognizing that content validity alone is insufficient for construct validation (Simms, 2008), we also conducted a one-tailed independent sample *t*-test to assess its predictive validity. The findings revealed a significantly higher perception of falling into a rabbit hole under the extreme content condition (M = 3.11, SD = 1.35) compared to the non-extreme content condition (M = 2.21, SD = 1.29), $t(389.89) = -6.73$, $p < .001$, $d = -0.68$. This provides criterion validity to our measure.

3.4.6. UI satisfaction

Based on the subscale of Overall Reactions to the Software from the Questionnaire for User Interface Satisfaction (Chin et al., 1988), we adopted six paired adjectives to evaluate the extent to which users were satisfied with the interface. These were (1) terrible/wonderful, (2) difficult/easy, (3) frustrating/satisfying, (4) inadequate power/adequate power, (5) dull/stimulating, and (6) rigid /flexible. We averaged them to form an index (M = 4.75, SD = 1.23), which was reliable, Cronbach's *alpha* = 0.91.

3.4.7. UX

UX was measured by four indicators adopted from Reynolds (2001), namely frustration (How frustrating was the experience on VIDNATION?), difficulty (How mentally difficult was it to use VIDNATION?), liking of the platform (Did you like using VIDNATION?), and interest in the platform (How interested are you in using VIDNATION again?). All questions were asked on a 7-point scale, ranging from 1 = not at all to 7 = very much.

3.4.8. Amount of prior online video viewing

Based on the Facebook addiction scale (Andreassen et al., 2012), we used six items to measure the amount of prior online video viewing. We asked participants "How often during the last year have you_?," with the following six statements: 1) spend a lot of thinking about online videos or planning to watch them; 2) feel an urge to watch videos online more and more; 3) watch online videos to forget about personal problems; 4) try to cut down on watching online videos without success; 5) become restless or troubled if I am prohibited from watching online videos; and 6) watch online videos so much that it has had a negative impact on my job/studies. We created an index by averaging the six items (M = 1.89, SD = 0.97, Cronbach's alpha = 0.91)

3.5. Construct validity

We tested the construct validity of these scales and measures based on the established criteria (Fornell and Larcker, 1981). Convergent validity is said to be achieved if the average variance extracted (AVE) is higher than 0.50 and the composite reliability (CR) is greater than 0.70. Furthermore, discriminant validity is demonstrated if the square root of AVE for each variable is greater than all the correlation coefficients involving that construct. As shown in Table 2, all constructs were valid based on these criteria.

4. Results

We first test manipulation effectiveness and then present the results for the research questions and hypotheses.

4.1. Manipulation check

After the interaction with the online video platform, we asked participants "Are you able to control the Next Video button?" with a binary response option Yes or No. Results from a Chi-square analysis indicated that most participants in the interpassive autoplay condition (n = 95) and manual play condition (n = 130) said Yes, whereas a majority of participants in the completely passive autoplay condition (n = 113) said No. The difference was statistically significant $\chi^2(2, N = 394) = 205.67$, $p < .001$.

Furthermore, we asked participants to indicate their level of agreement with the following two statements: "The platform automatically played the next video for me" and "The platform allowed me to turn the autoplay feature on and off" to measure perceived automation and perceived user control, respectively. Rated on a 7-point Likert scale (1 = strongly disagree and 7 = strongly agree), participants perceived a higher level of automation in the interpassive autoplay (M = 6.09, SD = 1.40) and completely passive autoplay condition (M = 6.45, SD = 1.15) compared to the manual play condition (M = 2.45, SD = 1.83), $F(2, 391) = 292.66$, $p < .001$, partial $\eta^2 = 0.60$. Furthermore, the perception of user control was the highest in the interpassive autoplay condition (M = 5.58, SD = 1.48), followed by manual play (M = 3.08, SD = 2.14) and completely passive autoplay condition (M = 2.64, SD = 1.95), $F(2, 391) = 92.06$, $p < .001$, partial $\eta^2 = 0.32$. The mean differences between interpassive autoplay and manual play, as well as between interpassive autoplay and completely passive autoplay, were significant at .001 level. Overall, the manipulation of autoplay modes was successful.

Regarding the manipulation effectiveness of content extremity,

Table 2
Convergent and discriminant validity of key study variables.

Variables	AVE	CR	1	2	3	4	5	6	7	8	9	10	11
Machine Heuristic (1)	.90	.97	.95[a]										
Control Heuristic (2)	.91	.97	.36***	.95[a]									
Inattentiveness (3)	.76	.93	.02	.06	.87[a]								
Expectancy Violation (4)	.91	.97	−0.29***	−0.15**	.52***	.95[a]							
Rabbit Hole Perception (5)	.69	.90	.33	.15**	.35***	.20***	.83[a]						
UI Satisfaction (6)	.70	.93	.47***	.45***	−0.25***	−0.56***	.17***	.84[a]					
Amount of Prior Online Video Viewing (7)	.69	.93	.40***	.20***	.37***	.21***	.40***	.23***	.83[a]				
Frustration (8)	n/a[b]	n/a[b]	−0.17***	−0.28***	.43***	.53***	.19***	−0.49***	.23***	n/a[b]			
Difficulty (9)	n/a[b]	n/a[b]	−0.03	−0.13**	.43***	.42***	.26***	−0.28***	.20***	.73***	n/a[b]		
Liking (10)	n/a[b]	n/a[b]	.61***	.45***	−0.21***	−0.56***	.21***	.83***	.25***	−0.45***	−0.23	n/a[b]	
Interest (11)	n/a[b]	n/a[b]	.61***	.42***	−0.18***	−0.53***	.23***	.79***	.29***	−0.37***	−0.16***	.88***	n/a[b]

[a] Bold values on the diagonal are the square root of AVE, an indicator of discriminant validity.
[b] Variables measured by a single item do not have AVE, CR, and the square root of AVE.

participants were asked how well (1 = not well at all and 7 = very well) the following adjectives describe the videos they saw on VIDNATION, including novel, new, odd, extreme, unrealistic, intense, niche, and hardcore. We created an index of perceived content extremity by averaging the eight items, which was reliable, Cronbach's alpha = 0.83. The index was used as the dependent variable in a one-tailed independent sample t-test. Results indicated that participants in the extreme condition (M = 3.10, SD = 1.15) perceived the video content to be significantly more extreme compared to their counterparts in the non-extreme condition (M = 2.59, SD = 1.09), $t(392) = 4.53$, $p < .001$, Cohen's d = 0.46. Thus, the manipulation of content extremity was successful as well.

4.2. Main effect of autoplay modes and content extremity

Given that video topic was used for stimulus sampling, it was added as a covariate in all analyses. An analysis of covariance (ANCOVA) showed that completely passive autoplay triggered the lowest control heuristic compared to interpassive autoplay and manual play. Similarly, completely passive autoplay led to the lowest UI satisfaction compared to the interpassive autoplay and manual play conditions. However, the difference between completely passive autoplay and manual play on UI satisfaction was only approaching significance. In addition, perceived frustration was significantly higher in the completely passive autoplay condition compared to the manual play and interpassive autoplay conditions. Regarding the interest in using the platform again, users showed much higher interest in the interpassive autoplay condition compared to the completely passive autoplay condition, but the difference was only marginally significant between the interpassive autoplay and the manual play conditions. We summarize the results in Table 3.

As shown in Table 4, ANCOVA results revealed that exposure to extreme content led to higher rabbit hole perception and interest in using the platform again compared to exposure to non-extreme video content.

Table 4
Main effect of content extremity.

Variables	Non-extreme	Extreme	Univariate F
Rabbit Hole Perception	2.22 (1.29)[a]	3.11 (1.35)[b]	$F(1, 391) = 44.83$, $p < .001$, partial $\eta^2 = 0.10$
Interest	4.00 (1.90)[a]	4.40 (1.96)[b]	$F(1, 391) = 4.07$, $p < .05$, partial $\eta^2 = 0.01$

Note. Standard deviation is in the parenthesis. Using LSD post hoc comparisons, means with no superscript in common differ at $p < .05$.

4.3. Model testing

We used Model 83 from PROCESS Macro (Hayes, 2017) to test the proposed moderated mediation model. Interpassive autoplay was used as the reference group in the indicator coding method used for dichotomizing the independent variable. The model was tested with 5,000 bootstrapped resamples and 95 % percentile confidence intervals (CIs).

4.3.1. Mediating roles of inattentiveness and negative expectancy violation

There was no significant interaction effect between autoplay modes and content extremity on inattentiveness, $F(2, 387) = 0.56$, $p = .57$. An examination of the moderated mediation model through both inattentiveness and negative expectancy violation showed that the model index was not significant for the comparison between manual play and interpassive autoplay, Index = 0.00, SE = 0.02, 95 % CI: [−0.02, 0.04], and for the comparison between completely passive autoplay and interpassive autoplay, Index = −0.00, SE = 0.01, 95 % CI: [−0.03, 0.02]. Thus, H1 was not supported.

We further tested whether the interaction effect on inattentiveness was dependent on the amount of prior online video viewing, by using a customized model based on Model 83 in PROCESS Macro. The three-way interaction on inattentiveness was significant, $F(2, 381) = 3.80$, $p < .05$. As shown in Fig. 3, interpassive autoplay (vs. manual play) increased

Table 3
Main effect of autoplay modes.

Variables	Interpassive Autoplay	Manual Play	Completely Passive Autoplay	Univariate F
Control Heuristic	4.88 (1.50)[a]	4.87 (1.41)[a]	2.56 (1.60)[b]	$F(2, 390) = 103.58$, $p < .001$, partial $\eta^2 = 0.35$
UI Satisfaction	5.00 (1.13)[a]	4.77 (1.22)[ab]	4.50 (1.29)[b]	$F(2, 390) = 5.61$, $p < .01$, partial $\eta^2 = 0.03$
Frustration	2.25 (1.69)[a]	2.52 (1.94)[ab]	3.03 (1.92)[b]	$F(2, 390) = 5.95$, $p < .01$, partial $\eta^2 = 0.03$
Interest	4.58 (1.87)[a]	4.11 (1.96)[ab]	3.93 (2.02)[b]	$F(2, 390) = 3.51$, $p < .05$, partial $\eta^2 = 0.02$

[a] Standard deviation is in the parenthesis. Using Fisher's Least Significance Difference (LSD) post hoc comparisons, means with no superscript in common differ at $p < .05$.

users' inattentiveness to the video content when the platform featured non-extreme content, whereas interpassive autoplay decreased inattentiveness when the platform showed extreme content. In both cases, the effect was true only for individuals who had low prior online video viewing.

Furthermore, results revealed that the conditional indirect effect on rabbit hole perception was true only through one mediator, i.e., inattentiveness, as shown in Fig. 4. A close examination of each condition indicated that manual play (vs. interpassive autoplay) decreased users' inattentiveness to the video content, $b = -0.90$, $se = 0.47$, $p = .056$, which further lowered rabbit hole perception only when the platform featured non-extreme content and only for those who had low prior online video viewing (see Table 5). Given that the findings are in a different direction than hypothesized, H4 was not supported.

4.3.2. *Mediating roles of machine heuristic and negative expectancy violation*

We used Model 83 to examine the interaction effect between autoplay modes and content extremity on rabbit hole perception through machine heuristic and negative expectancy violation. Again, there was no significant interaction effect on the triggered machine heuristic. The index of the moderated mediation model was not significant either. Thus, H2 was not supported. Adding the amount of prior online video viewing as a second moderator, moderating the conditional effect of content extremity on the relationship between autoplay modes and machine heuristic, did not result in a significant model either. Therefore, H5 was not supported.

4.3.3. *Mediating roles of control heuristic and negative expectancy violation*

Regarding the mediating role of control heuristic, results from Model 83 revealed a significant index of moderated mediation when comparing interpassive autoplay with completely passive autoplay, Index $= -0.03$, $SE = 0.02$, 95 % CI: $[-0.07, -0.00]$. A close examination of each condition showed that interpassive autoplay triggered the control heuristic to a greater extent compared to completely passive autoplay, and the invoked control heuristic was associated with less negative expectancy violation, which further lowered rabbit hole perception. The serial mediation model comparing interpassive autoplay and completely passive autoplay was true regardless of content extremity, as shown in Table 6. Given that we hypothesized that interpassive autoplay (vs. completely passive autoplay) would increase the control heuristic under the extreme content condition only, H3 was partially supported. We present the moderated mediation model in Fig. 5.

In addition, results from Model 83 suggested that the path from the control heuristic to rabbit hole perception was significant. This implies that the control heuristic that is not associated with negative expectancy violation is positively related to rabbit hole perception.

4.4. *Summary of the findings*

This study found that interpassive autoplay was favored by users as it was more likely to trigger the control heuristic, led to higher UI satisfaction and interest in using the platform, and caused less frustration, compared to manual play and completely passive autoplay. We found that the invoked control heuristic from interpassive autoplay was associated with higher rabbit hole perception than completely passive autoplay. However, when the triggered control heuristic led to less negative expectancy violation, it was associated with a lower rabbit hole perception. Compared to manual play, interpassive autoplay increased users' inattentiveness to the video content, which further increased rabbit hole perception when the platform showed non-extreme content, but this effect was true only for those who scored low in the amount of prior online video viewing.

5. Discussion

This study compares the UX and UI satisfaction of three modes of autoplay based on the affordance of interpassivity. Furthermore, it reveals the conditions and psychological mechanisms that drive the effect of interpassive autoplay on rabbit hole perception.

5.1. *Is interpassivity a psychologically valid concept?*

The first goal of our study was to compare three modes of autoplay based on the affordance of interpassivity. Ontologically, interpassivity is the action possibility afforded by autoplay, as it allows users to enjoy video automatically played one after another, but also allows users to toggle the feature off if they so desire. However, whether users perceive both automation and interactivity from autoplay is under-explored. This

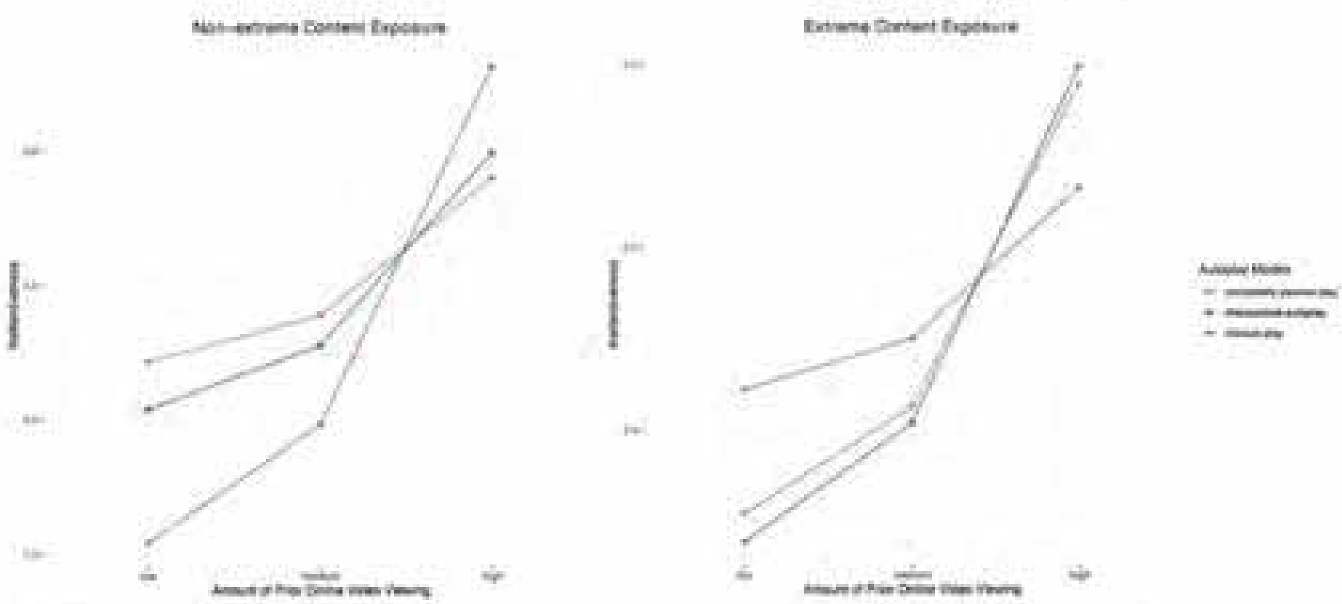
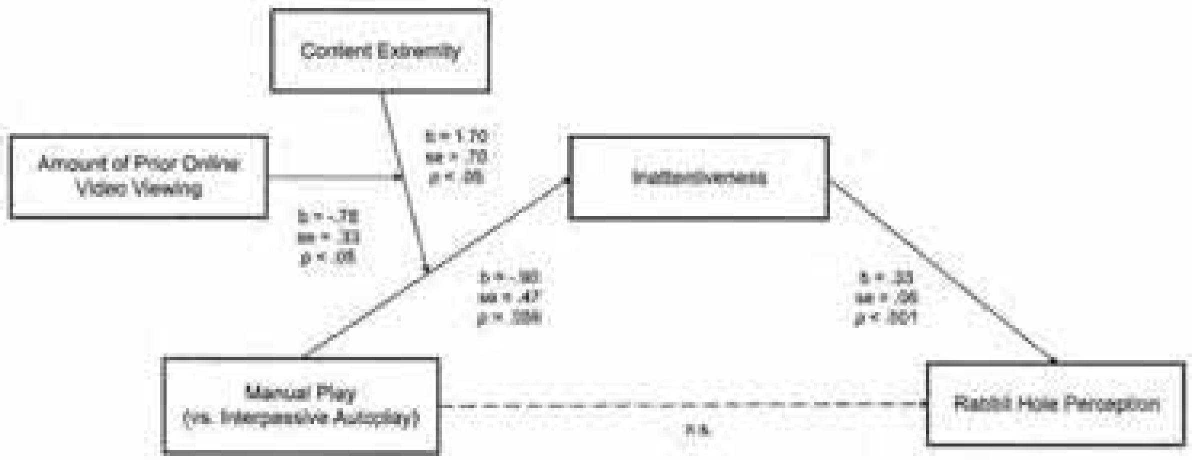


Fig. 3. Conditional effect of amount of prior online video viewing on the effect of autoplay modes and content extremity on inattentiveness under both non-extreme (left) and extreme (right) content exposure.



Note. Manual play was coded as 1 and interpassive autoplay was coded as 0.

Fig. 4. Conditional moderated mediation model on rabbit hole perception through inattentiveness.

study fills that gap by examining the psychological effect of interpassivity. We found that users perceive both high automation and interactivity from interpassive autoplay compared to manual play and completely passive autoplay. Therefore, we argue that interpassivity is an ontologically and psychologically valid experience, which is afforded by autoplay.

Furthermore, the two unique affordances of autoplay, i.e., automation and interactivity, help us understand how the three autoplay modes differ from the users' perspective. We found that manual play results in high perceived interactivity but low perceived automation, whereas completely passive autoplay leads to high perceived automation but low perceived interactivity. Interpassive autoplay is the only mode that causes both high perceived automation and interactivity. This co-occurrence means that perceived automation and interactivity are not at odds with each other. Users can achieve both from different aspects of autoplay. Given that automation is primarily system-driven and interactivity is mostly user-driven, the collaboration between machines and humans seems to result in better user experience, characterized as interpassivity. Previous studies have shown that the synergy between machine agency and human agency tends to generate better interaction outcomes in the context of human-computer interaction (Kang and Lou, 2022; Sundar, 2020).

5.2. *Do users like the autoplay feature?*

We found that users have positive perceptions of interpassive autoplay, as it triggers the control heuristic, making users believe that they have a higher sense of control over their viewing experience. Additionally, the presence of interpassive autoplay leads to a more favorable reaction to the online video platform as a whole and increases users' intentions to continue using it. More interestingly, interpassive autoplay seems to lower users' frustration during their interaction with the platform compared to completely passive autoplay. The positive evaluation of interpassive autoplay echoes a recent study on the uses and gratifications of automated features, which found that users enjoy autoplay because it provides convenience, user control, and user profiling (Chen et al., 2023). Among the three gratifications, the important role of user control is confirmed and highlighted by the current study. The positive correlations between the control heuristic and UX and UI satisfaction in the study further explain why users enjoy autoplay as an interface feature.

Our findings show that the triggered control heuristic is higher in interpassive autoplay and manual play compared to completely passive autoplay. This suggests that the affordance of interactivity seems to give rise to user control (Molina and Sundar, 2022; Sun and Sundar, 2022). Considering that both interpassive autoplay and manual play featured the Next Video button, it appears that being able to assert one's agency during the playing of the videos is an important aspect that affects users' sense of control. Previous research has also pointed out that users appreciate the opportunity to reclaim their agency rather than being passive recipients of the recommendations (Lukoff et al., 2021; Roffarello and De Russis, 2022).

Furthermore, we found that there is no difference between interpassive autoplay and manual play in terms of the activated control heuristic; this suggests that the simple toggle option is sufficient to provide users with the same level of control during autoplay reception as having the ability to manually control each and every video in the non-autoplay scenario. This finding indicates the power of the toggle button, a point further discussed in relation to rabbit hole perception.

5.3. *Do users like extreme content?*

In addition, we found that users seem to enjoy extreme content. When the platform features extreme video content, it leads to higher liking of the platform and greater interest in using the platform again. Given that extreme video is by nature more interesting, rare, and novel than non-extreme video content, the positive evaluation toward the online video platform is not surprising. Previous research has also demonstrated the linkage between perceived enjoyment of the content and positive attitudes toward the system (Praveena and Thomas, 2014).

However, the preference for extreme videos should be interpreted by keeping in mind our operationalization of content extremity. We defined content extremity as daily-life activities that escalate to an extreme. This definition suggests that the majority of recommended video content is socially desirable, and is not focused on harmful and false information. Thus, our findings indicating that users tend to enjoy daily-life video content that goes to an extreme do not necessarily generalize to less desirable content such as radical political speech and misinformation.

5.4. *Can interpassive autoplay increase rabbit hole perception when content becomes extreme?*

One major finding of the study is that interpassive autoplay can make a difference in users' rabbit hole perception. Two mediators can explain the effect of interpassive autoplay on rabbit hole perception under both extreme and non-extreme content conditions. One is the control

Table 5
Conditional moderated mediation effect of modes of autoplay on rabbit hole perception through inattentiveness under both extreme and non-extreme content.

Modes of Autoplay	Content Extremity	Amount of Prior Online Video Viewing	Effect (B)	BootSE	95 % CI
Indirect Effect: Autoplay Mode → Inattentiveness → Rabbit Hole Perception					
Manual play (vs. interpassive autoplay)	Non-extreme	Low	−0.16	.08	[−0.34, −0.00]
Manual play (vs. interpassive autoplay)	Non-extreme	Medium	−0.10	.07	[−0.24, 0.03]
Manual play (vs. interpassive autoplay)	Non-extreme	High	.10	.14	[−0.19, 0.38]
Manual play (vs. interpassive autoplay)	Extreme	Low	.14	.12	[−0.08, 0.38]
Manual play (vs. interpassive autoplay)	Extreme	Medium	.06	.08	[−0.09, 0.25]
Manual play (vs. interpassive autoplay)	Extreme	High	−0.11	.15	[−0.09, 0.25]
Indices of Conditional Moderated Mediation:					
Low use: Index = 0.30, BootSE = 0.15, 95 % CI: [0.04, 0.61]					
Medium use: Index = 0.17, BootSE = 0.11, 95 % CI: [−0.03, 0.41]					
High use: Index = −0.21, BootSE = 0.21, 95 % CI: [−0.63, 0.19]					
Completely passive autoplay (vs. interpassive autoplay)	Non-extreme	Low	.06	.11	[−0.16, 0.25]
Completely passive autoplay (vs. interpassive autoplay)	Non-extreme	Medium	.04	.08	[−0.13, 0.18]
Completely passive autoplay (vs. interpassive autoplay)	Non-extreme	High	−0.03	.17	[−0.38, 0.32]
Completely passive autoplay (vs. interpassive autoplay)	Extreme	Low	.03	.09	[−0.15, 0.21]
Completely passive autoplay (vs. interpassive autoplay)	Extreme	Medium	.01	.07	[−0.13, 0.15]
Completely passive autoplay (vs. interpassive autoplay)	Extreme	High	−0.02	.13	[−0.27, 0.20]
Indices of Conditional Moderated Mediation:					
Low use: Index = −0.03, BootSE = 0.14, 95 % CI: [−0.29, 0.27]					
Medium use: Index = −0.02, BootSE = 0.11, 95 % CI: [−0.22, 0.20]					
High use: Index = −0.01, BootSE = 0.22, 95 % CI: [−0.43, 0.45]					

Table 6
Conditional indirect effect of autoplay modes on rabbit hole perception through the control heuristic and negative expectancy violation.

Autoplay Modes	Content Extremity	Effect (B)	BootSE	95 % CI
Indirect Effect: Autoplay Modes → Control Heuristic → Negative Expectancy Violation → Rabbit Hole Perception				
Manual play (vs. interpassive autoplay)	Non-extreme	.00	.01	[−0.01, 0.02]
Manual play (vs. interpassive autoplay)	Extreme	−0.00	.01	[−0.02, 0.02]
Index of Moderated Mediation: Index = −0.00, BootSE = 0.01, 95 % CI: [−0.03, 0.02]				
Completely passive autoplay (vs. interpassive autoplay)	Non-extreme	.09	.04	[.03, 0.17]
Completely passive autoplay (vs. interpassive autoplay)	Extreme	.06	.03	[.02, 0.11]
Index of Moderated Mediation: Index = −0.03, BootSE = 0.02, 95 % CI: [−0.07, −0.00]				

heuristic. We found that interpassive autoplay is more likely to trigger the control heuristic, which is associated with lower negative expectancy violation and rabbit hole perception compared to completely passive autoplay. This finding indicates that when users feel agentic, probably due to the interactivity aspect of autoplay, it makes them more likely to believe that the recommendation meets their expectations, which further lowers their awareness that they are being led down a rabbit hole.

While we proposed that the effect of interpassive autoplay on rabbit hole perception would go through two mediators, we found that the invocation of the control heuristic tends to increase rabbit hole perception when it does not mitigate negative expectancy violation. This result highlights that the quality of the control heuristic activated by the interpassive autoplay matters. If the control heuristic can increase users' consciousness of their media use, it leads to higher rabbit hole perception. By contrast, if the control heuristic engenders complacency by lulling them to believe that the videos recommended align with their expectations, then it tends to be associated with a lower rabbit hole perception.

Considering that the control heuristic is mostly from the interactivity aspect of autoplay, it seems likely that the presence of the Next Video button and the toggle option contribute to the higher control heuristic activated in the interpassive autoplay condition compared to the completely passive autoplay condition. Given that we did not find any difference in the activated control heuristic betweeninterpassive autoplay and manual play, it implies that the Next Video button is probably not the primary driver of the control heuristic. Instead, the toggle option seems to play a critical role in giving rise to a greater sense of user control and further influences user experience and perceptions of extreme content in the interpassive autoplay condition.

It is worth noting that the toggle option offers both modality interactivity (i.e., allowing users to turn the feature on and off through a toggle button) and source interactivity (i.e., allowing users to see themselves as the source of the recommendation). If users are influenced by modality interactivity, the presence of a unique interaction technique, i.e., the toggle option, may increase the fun and pleasure of using the autoplay feature. However, a playful user experience is also found to activate heuristic information processing, leading to fewer message-related thoughts (Oh and Sundar, 2015). In light of this, it is possible that the playfulness engendered by the toggle option triggers a false sense of user control, with users overlooking the negatively violated experience and thereby becoming less aware of the rabbit hole. If users are primarily driven by source interactivity afforded by the toggle option, they are likely to see themselves as the source of the recommendation, which increases their sense of control. However, being the gatekeeper of information also demands users' cognitive resources (Kang and Sundar, 2013). The depleted ego may hamper the evaluation of their negatively violated experience, leading to less rabbit hole perception. Together, the sense of user control afforded by interactivity could be a double-edged sword. High-quality user control could enhance users' cognition and increase rabbit hole perception. However, when the triggered control heuristic is accompanied by perceived playfulness or ego depletion, it may undermine users' cognitive engagement with the content, leading them to believe that what the system recommends matches one's expectation, thus lowering their rabbit hole perception.

Aside from the control heuristic, inattentiveness is another mechanism that accounts for the effect of autoplay on rabbit hole perception.

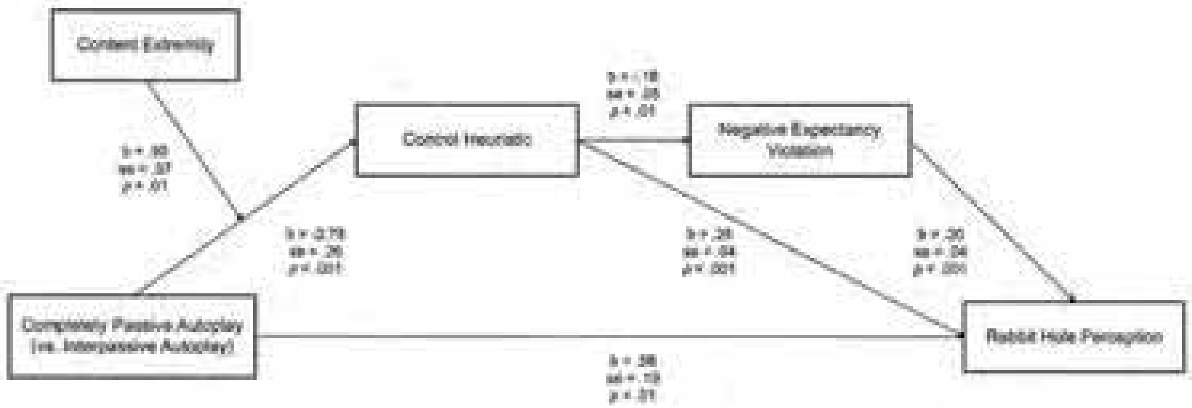


Fig. 5. Testing moderated mediation model on rabbit hole perception through the control heuristic and negative expectancy violation.

We found that interpassive autoplay tends to cause higher inattentiveness compared to manual play when the platform features non-extreme content, and the perceived inattentiveness in turn leads to higher rabbit hole perception. This finding seems contradictory given that there is no rabbit hole when the recommended content is not extreme. Considering that rabbit hole perception is broadly defined as users' sensitivity to content extremity, it means that users are more sensitive to the tendency to go down the rabbit hole when the platform features common, mainstream, and mundane content. Given the prevalence of extreme content in online video platforms, such as YouTube (Brown et al., 2022; Tang et al., 2021; Tufekci, 2018), users probably expect extreme content to be upcoming even while being fed with non-extreme content, thus increasing their rabbit hole perception.

However, the positive relationship between inattentiveness and rabbit hole perception does not align with our hypothesis, which predicted that users would experience mental underload when automation takes effect, thus reducing rabbit hole perception (Young and Stanton, 2002). The discrepancy is probably due to different conceptualizations of inattentiveness. If we interpret inattentiveness as a lack of interest in the video content played on the platform, then the positive correlation between inattentiveness and rabbit hole perception makes sense, as it suggests that if users are less involved in the videos to which they are exposed, they may have more attentional resources to evaluate the nature of the recommendations as a whole, resulting in higher rabbit hole perception. Our findings reveal that this indirect effect is true only for those who are not heavy online video viewers. It suggests that users with higher self-regulation are more inattentive to the content when the platform features non-extreme content under autoplay. It is probably because users are not motivated to process the recommended video when they have a viewing goal to begin with. Such inattentiveness seems to trigger higher rabbit hole perception among non-heavy online video viewers. Again, negative expectancy violation did not play a role in the model; instead, inattentiveness is a powerful factor driving rabbit hole perception. It means that negative expectancy violation may not necessarily precede rabbit hole perception when users are mindless during their video reception.

One unexpected finding is that the machine heuristic did not mediate the effect of autoplay on rabbit hole perception under both extreme and non-extreme content conditions. Given that machine heuristic is often triggered when the source of interaction is a machine (Molina and Sundar, 2022; Sundar and Kim, 2019), one plausible explanation is that the machine heuristic is not sufficiently invoked as the machine-ness cues are very subtle on the current interface. Considering that we did not mention words such as AI and algorithms when introducing the platform to users, cues pertaining to machine-ness do not seem to be available in users' minds, thus undermining the invocation of the machine heuristic. Furthermore, online video platforms like YouTube may be considered an older medium compared to emerging and newer media technologies, such as ChatGPT. Again, the lack of machine-ness cues, suggesting the machine is intelligent, competent, and cool, may have hindered the activation of the machine heuristic.

Overall, we found that autoplay can increase users' rabbit hole perception through two mediators: the control heuristic and inattentiveness. The former is true regardless of content extremity, while the latter works only when the platform features non-extreme content and is for light online video viewers. By revealing the mechanisms and conditions under which interpassive autoplay can influence users' awareness of extreme content, this study provides ideas for the socially responsible design of autoplay, as discussed below.

5.5. Practical implications

First, our manipulation of three modes of autoplay was successful, triggering different levels of perceived automation and interactivity. The success was built upon several failures in the pre-test as simply presenting the autoplay toggle option and the 3-second countdown timer did not help users fully realize the affordance of interpassivity. One solution that worked in the present study is the tour, in which we introduced the major features of the platform prior to video playing. The tour seems to increase users' understanding of the system's functionality in general and the autoplay feature in particular. Given that the design of autoplay in platforms such as YouTube is very subtle, designers should consider adding a tour or tooltip to help users better understand how autoplay works.

Second, the positive evaluation of extreme content may suggest that designers incorporate more novel, rare, and unusual content in the recommendation algorithms. However, it is important to consider the nature of the extreme content. Designers should be aware that some extreme content could be harmful as it contains misinformation (Tang et al., 2021), radicalization (Ledwich and Zaitsev, 2019), and sexually suggestive material (Kaiser, 2019). It is therefore irresponsible and unethical to feed users extreme videos to increase positive perceptions of the platform.

Third, interpassive autoplay seems to promote users' rabbit hole perception through two mechanisms: the control heuristic and inattentiveness. Given that the control heuristic is primarily triggered by interactivity and inattentiveness is mostly caused by the automation aspect of autoplay, featuring either user control or the allure of convenience and relaxation is promising to activate the corresponding psychological mechanism that drives rabbit hole perception. One risk pertaining to the affordance of interactivity is that the invoked control heuristic may lower rabbit hole perception if the negative expectancy

violation is attenuated. As such, designers should be cautious of the quality of user control triggered by the autoplay toggle option. The kind of user control that empowers users and raises their consciousness seems to enhance rabbit hole perception, but the part that makes them feel complacent seems to make them overlook the negatively violated experience, thus resulting in lesser rabbit hole perception. Designers should consider measuring the level of the activated user control during their interaction with the toggle option and come up with intervention strategies to promote more mindful and conscious use of the autoplay feature.

5.6. *Limitations and future studies*

This study has some methodological limitations that need to be addressed in future studies. We manipulated the completely passive autoplay condition by disabling several control buttons on the platform, including play, pause, and volume control. While it simulates the scenario in which users lean back and enjoy the play of the videos without active participation, this setup is not natural as users are forced to watch the videos without considering their preferences. This design may have negatively influenced user experience. Future studies should observe how completely passive autoplay influences rabbit hole perception in a natural environment. This effort will help validate the findings of the current study.

The second limitation pertains to the lack of behavioral data. We are unsure how users interact with the autoplay feature, such as how many participants paid attention to the toggle option and how many tried to turn the autoplay feature off. The sheer presence of the autoplay feature only allows us to observe the cue effect of the interactivity affordance (Sundar, 2020). While revealing the relationship between interactivity cues and the control heuristic is informative, it would be more interesting to explore the action effect, i.e., how the actual interaction with the autoplay toggle button influences rabbit hole perception.

Third, the current study mimics YouTube's autoplay design, which may not be representative of the autoplay features on other platforms. For example, YouTube allows users to turn the autoplay feature on and off through the toggle option, but the option to disable autoplay on other platforms is often hidden or non-existent (Roffarello et al., 2023). This deceptive design may prevent users from experiencing the interactivity aspect of the autoplay feature, thereby reducing the activated control heuristic and leading to a lower perception of falling into a rabbit hole. Thus, it is crucial to validate the study's findings across various platforms using the autoplay feature and to explore how the placement of the autoplay button influences perceived user control over the feature and their rabbit hole perception.

Fouth, perceived content extremity was below the midpoint, as the study focused on routine daily life activities taken to an extreme. It is likely that a study involving hot button political issues, health misinformation, hate speech, and other socially undesirable media content would result in higher perceived content extremity and a better observation of rabbit hole perception. However, we are mindful of the ethicality of conducting such research with human subjects and therefore decided to avoid exposing our participants to problematic media content in the interest of protecting their well-being.

Last, the recommendation algorithms of YouTube are evolving. A recent study found that they rarely send people down the rabbit hole (A. Y. Chen et al., 2023). However, if users are actively seeking extreme content, the algorithms may learn their preferences and increase exposure to content from alternative and extremist channels. Future studies would do well to explore whether user-initiated exposure to extreme content can influence rabbit hole perception under different modes of autoplay.

6. Conclusion

Based on the affordance of interpassivity, this study compared three modes of autoplay and examined their relative influences on rabbit hole perception when individuals are shown progressively extreme content. Despite the sweeping criticism of autoplay for undermining users' sense of control (Lukoff et al., 2021, 2023) and leading to problematic use (Roffarello and De Russis, 2022; Schaffner et al., 2023), we found that interpassive autoplay is generally appreciated as an interface feature and can promote users' rabbit hole perception under certain conditions. Future studies should examine the dynamics of autoplay use. Currently, autoplay is a conditionally autonomous feature that relies on user control. A more advanced and highly automated autoplay could identify users' cognition, affect, and behaviors and make corresponding adjustments in real time. In this way, findings from the current study can be used in a theoretically and empirically grounded manner to design future variations of autoplay that can serve users' needs while also protecting their welfare.

CRediT authorship contribution statement

Cheng Chen: Conceptualization, Data curation, Formal analysis, Methodology, Writing – original draft, Writing – review & editing. **Jingshi Kang:** Conceptualization, Methodology, Writing – review & editing. **Pejman Sajjadi:** Conceptualization, Software, Writing – review & editing. **S. Shyam Sundar:** Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

10 Most Watched Categories of YouTube Videos. (2022, February 14). Mogall. https://mogall.com/blog/10-most-watched-categories-of-youtube-videos.

Andreassen, C.S., Torsheim, T., Brunborg, G.S., Pallesen, S., 2012. Development of a facebook addiction scale. Psychol. Rep. 110 (2), 501–517. https://doi.org/10.2466/02.09.18.PR0.110.2.501-517.

Brown, M., Bisbee, J., Lai, A., Bonneau, R., Nagler, J., Tucker, J.A., 2022. Echo chambers, rabbit holes, and algorithmic bias: how YouTube recommends content to real users. SSRN Electr. J. https://doi.org/10.2139/ssrn.4114905.

Carnagey, N.L., Anderson, C.A., Bushman, B.J., 2007. The effect of video game violence on physiological desensitization to real-life violence. J. Exp. Soc. Psychol 43 (3), 489–496. https://doi.org/10.1016/j.jesp.2006.05.003.

Chen, A.Y., Nyhan, B., Reifler, J., Robertson, R.E., Wilson, C., 2023a. Subscriptions and external links help drive resentful users to alternative and extremist YouTube channels. Sci. Adv 9 (35), eadd8080. https://doi.org/10.1126/sciadv.add8080.

Chen, C., Lee, S., Sundar, S.S., 2023b. Interpassivity instead of interactivity? The uses and gratifications of automated features. Behav. Inf. Technol. https://doi.org/10.1080/0144929X.2023.2184174.

Chin, J.P., Diehl, V.A., Norman, K.L., 1988. Development of an instrument measuring user satisfaction of the human-computer interface. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 213–218. https://doi.org/10.1145/57167.57203.

Fornell, C., Larcker, D.F., 1981. Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research 18 (1), 39–50. https://doi.org/10.2307/3151312.

Gerbner, G., 1967. An institutional approach to mass communications research. In: Communication theory and research: Proceedings of the First International Symposium. Charles C Thomas, pp. 429–445.

Gerbner, G., 1969. Toward 'cultural indicators': the analysis of mass mediated public message systems. The Analysis of Communication Content: Developments in Scientific Theories and Computer Techniques. John Wiley & Sons, pp. 123–132.

Hassenzahl, M., Tractinsky, N., 2006. User experience-a research agenda. Behav. Inf. Technol 25 (2), 91–97. https://doi.org/10.1080/01449290500330331.

Hayes, A.F., 2017. Introduction to mediation, moderation, and Conditional Process analysis: A regression-Based Approach. Guilford publications.

Kaiser, J., 2019. The implications of venturing down the rabbit hole. Inter. Policy Rev. https://policyreview.info/articles/news/implications-venturing-down-rabbit-hole/1406.

Kang, H., Lou, C., 2022. AI agency vs. human agency: understanding human–AI interactions on TikTok and their implications for user engagement. J. Comp.-Med. Commun. 27 (5) https://doi.org/10.1093/jcmc/zmac014 zmac014.
Kang, H., Sundar, S.S., 2013. Depleted egos and affirmed selves: the two faces of customization. Comput. Human Behav. 29 (6), 2273–2280. https://doi.org/10.1016/j.chb.2013.05.018.
LaRose, R., Lin, C.A., Eastin, M.S., 2003. Unregulated internet usage: addiction, habit, or deficient self-regulation? Media Psychol 5 (3), 225–253. https://doi.org/10.1207/S1532785XMEP0503_01.
Ledwich, M., & Zaitsev, A. (2019). Algorithmic extremism: examining YouTube's rabbit hole of radicalization (arXiv:1912.11211). arXiv. http://arxiv.org/abs/1912.11211.
Livingstone, S., Kirwil, L., Ponte, C., Staksrud, E., 2014. In their own words: what bothers children online? Europ. J. Commun. 29 (3), 271–288. https://doi.org/10.1177/0267323114521045.
Lukoff, K., Lyngs, U., Shirokova, K., Rao, R., Tian, L., Zade, H., Munson, S.A., Hiniker, A., 2023. SwitchTube: a proof-of-concept system introducing "adaptable commitment interfaces" as a tool for digital wellbeing. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, pp. 1–22. https://doi.org/10.1145/3544548.3580703.
Lukoff, K., Lyngs, U., Zade, H., Liao, J.V., Choi, J., Fan, K., Munson, S.A., Hiniker, A., 2021. How the design of YouTube influences user sense of agency. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, pp. 1–17. https://doi.org/10.1145/3411764.3445467.
Markmann, S., Grimme, C., 2021. Is YouTube still a radicalizer? An exploratory study on autoplay and recommendation. Disinfor. Open Online Media 3, 50–65. https://doi.org/10.1007/978-3-030-87031-7_4.
McCrosky, J., & Geurkink, B. (2021). YouTube regrets: a crowdsourced investigation into YouTube's recommendation algorithm. https://foundation.mozilla.org/en/youtube/findings/.
Miller, D.T., 1976. Ego involvement and attributions for success and failure. J. Pers. Soc. Psychol. 34 (5), 901–906. https://doi.org/10.1037/0022-3514.34.5.901.
Molina, M.D., Sundar, S.S., 2022. When AI moderates online content: effects of human collaboration and interactive transparency on user trust. J. Comp.-Med. Commun. 27 (4) https://doi.org/10.1093/jcmc/zmac010.
Nass, C., Mason, L., 1990. On the study of technology and task: a variable-based approach. Organizations and Communication Technology. Sage publications. https://doi.org/10.4135/9781483325385.n3.
Oh, J., Sundar, S.S., 2015. How does interactivity persuade? An experimental test of interactivity on cognitive absorption, elaboration, and attitudes: persuasive effects of interactivity. J. Commun. 65 (2), 213–236. https://doi.org/10.1111/jcom.12147.
Pfeifer, R., 1996. Um Die Echo Gehuche. Echer 41, 71.
Praveena, K., Thomas, S., 2014. Continuance intention to use facebook: a study of perceived enjoyment and TAM. Bonfring Int. J. Indus. Engineer. Manag. Sci. 4 (1), 24–29. https://doi.org/10.9756/BIJIEMS.4794.
Ramirez Jr, A., Wang, Z., 2008. When online meets offline: an expectancy violations theory perspective on modality switching. J. Commun. 58 (1), 20–39. https://doi.org/10.1111/j.1460-2466.2007.00372.x.
Reynolds, C.J. (2001). The sensing and measurement of frustration with computers [Massachusetts Institute of Technology]. http://hdl.handle.net/1721.1/39579.
Ribeiro, M.H., Ottoni, R., West, R., Almeida, V.A.F., Meira, W., 2020. Auditing radicalization pathways on YouTube. In: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, pp. 131–141. https://doi.org/10.1145/3351095.3372879.
Roffarello, A.M., De Russis, L., 2022. Towards understanding the dark patterns that steal our attention. In: CHI Conference on Human Factors in Computing Systems Extended Abstracts, pp. 1–7. https://doi.org/10.1145/3491101.3519829.
Roffarello, A.M., Lukoff, K., De Russis, L., 2023. Defining and identifying attention capture deceptive designs in digital interfaces. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, pp. 1–19. https://doi.org/10.1145/3544548.3580729.
Schaffner, B., Stefanescu, A., Campili, O., Chetty, M., 2023. Don't let Netflix drive the bus: user's sense of agency over time and content choice on Netflix. Proc. ACM Human-Comp. Inter 7 (CSCW1), 1–32. https://doi.org/10.1145/3579604.
Smith, E.G., 1998. The construct of content validity. Soc. Indic. Res. 45, 83–117. https://doi.org/10.1023/A:1006985528729.
Stöcker, C., Preuss, M., 2020. Riding the wave of misclassification: how we end up with extreme YouTube content. In: International Conference on Human-Computer Interaction, pp. 359–375. https://doi.org/10.1007/978-3-030-49570-1_25.
Sun, Y., Sundar, S.S., 2022. Exploring the effects of interactive dialogue in improving user control for explainable online symptom checkers. In: CHI Conference on Human Factors in Computing Systems Extended Abstracts, pp. 1–7. https://doi.org/10.1145/3491101.3519668.
Sundar, S.S., 2008. The MAIN model: a heuristic approach to understanding technology effects on credibility. Digital Media, Youth, and Credibility. The MIT Press, pp. 73–100. https://doi.org/10.1162/dmal.9780262562324.073.
Sundar, S.S., 2020. Rise of machine agency: a framework for studying the psychology of human–AI interaction (HAII). J. Comp. Med. Commun. 25 (1), 74–88. https://doi.org/10.1093/jcmc/zmz026.
Sundar, S.S., Kim, J., 2019. Machine heuristic: when we trust computers more than humans with our personal information. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1–9. https://doi.org/10.1145/3290605.3300768.
Sundar, S.S., Kim, J., Rosson, M.B., Molina, M.D., 2020. Online privacy heuristics that predict information disclosure. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–12. https://doi.org/10.1145/3313831.3376854.
Tang, L., Fujimoto, K., Amith, M.(Tuan), Cunningham, R, Costantini, R.A. York, F, Xiong, G. Boom, J.A. Tao, C, 2021. Down the Rabbit Hole" of vaccine misinformation on YouTube: network exposure study. J. Med. Internet Res. 23 (1), e23262. https://doi.org/10.2196/23262.
Tufekci, Z. (2018). YouTube, the great radicalizer. The New York Times. https://onlinebr.unesco.fr/teaching/2018/03/YouTube.pdf.
Waddell, T.F., 2018. A robot wrote this? How perceived machine authorship affects news credibility. Digital Journal. 6 (2), 236–255. https://doi.org/10.1080/21670811.2017.1384319.
Whiteside, J., Wixon, D., 1987. The dialectic of usability engineering. Human-Comp. Inter.-INTERACT '87, 17–20. https://doi.org/10.1016/B978-0-444-70304-0.50011-3.
Woolley, K., Sharif, M.A., 2022. Down a rabbit hole: how prior media consumption shapes subsequent media consumption. J. Mark. Res. 59 (3), 453–471. https://doi.org/10.1177/00222437211055403.
Yang, H., & Sundar, S.S. (2020, May 21). Machine heuristic: a concept explication and development of a scale. The 70th annual conference of the International Communication Association, Virtual Event USA.
Young, M.S., Stanton, N.A., 2002. Malleable attentional resources theory: a new explanation for the effects of mental underload on performance. Hum. Factors 44 (3), 365–375. https://doi.org/10.1518/0018720024497709.
Zitek, E., 1997. The plague of fantasies. Verso.

3

healthcare

MDPI

Systematic Review

The Effects of Social Feedback Through the "Like" Feature on Brain Activity: A Systematic Review

Artemisa R. Dores [1,2,*], Miguel Peixoto [1,2], Carina Fernandes [2,3,4], António Marques [1]

Likes

and Fernando Barbosa [2]

1 Psychosocial Rehabilitation Laboratory, Center for Rehabilitation Research (LabRP-CIR), Escola Superior de Saúde, Instituto Politécnico do Porto, 4200-072 Porto, Portugal; mpe@ess.ipp.pt (M.P.); ajmarques@ess.ipp.pt (A.M.)
2 Laboratory of Neuropsychophysiology, Faculty of Psychology and Education Sciences, University of Porto, 4200-135 Porto, Portugal; carinafernandes@ufp.edu.pt (C.F.); fbarbosa@fpce.up.pt (F.B.)
3 Faculty of Human and Social Sciences, University Fernando Pessoa, 4249-004 Porto, Portugal
4 Molecular Oncology and Viral Pathology Group, Research Center of IPO (CI-IPOP) & RISE@CI-IPOP (Health Research Network), Portuguese Oncology Institute of Porto (IPO Porto)/Porto Comprehensive Cancer Center (Porto.CCC), 4200-072 Porto, Portugal
* Correspondence: artemisa@ess.ipp.pt

Likes

ial media (SM) use is a growing concern, particularly
s these platforms for social interactions important
is characterized by excessive, uncontrolled usage
fessional aspects. Despite the ongoing debate over
diagnostic category, the impact of social feedback,
on brain activity remains under scrutiny. Objective:
he neural correlates of online social feedback, focus-
k on brain activity using fMRI and EEG. Methods:
ions of the Preferred Reporting Items for Systematic
(PRISMA). Results: The review included 11 studies
rain structures such as the amygdala, ventromedial
l striatum involved in reward processing. Positive
he nucleus accumbens (NACC), vmPFC, and amyg-
eased SM use intensity. Negative feedback activates
PFC) and left medial prefrontal cortex (mPFC). Be-
eedback influences subsequent social interactions.
lisparities in the literature regarding the neural response to social feedback, emphasizing the need for further research to clarify the roles of sex, personality traits, and the person giving feedback. Overall, understanding the neurobiological underpinnings of SM engagement is essential for developing effective interventions to prevent or address the negative effects of excessive SM use.

check for updates

Academic Editor: D

Received: 19 Octobe
Revised: 12 Decemb
Accepted: 30 December 2024
Published: 4 January 2025

Citation: Dores, A.R.; Peixoto, M.; Fernandes, C.; Marques, A.; Barbosa, F. The Effects of Social Feedback Through the "Like" Feature on Brain Activity: A Systematic Review. *Healthcare* **2025**, *13*, 89. https://doi.org/10.3390/healthcare13010089

Keywords: social media; like button; EEG; fMRI

1. Introduction

Problematic use of social media (SM) platforms is an ever-growing concern that affects various populations, especially people from low-income countries [1], with moderate to low school achievement, low parental control [2], low self-esteem [3], feelings of loneliness, fear of negative evaluation, and behavioral inhibition [4]. Adolescents seem to be amongst the more vulnerable populations due to SMs' unique appeal to their age range. They are

healthcare

MDPI

Systematic Review

The Effects of Social Feedback Through the "Like" Feature on Brain Activity: A Systematic Review

Artemisa R. Dores [1,2,*], Miguel Peixoto [1,2], Carina Fernandes [2,3,4], António Marques [1] and Fernando Barbosa [2]

1 Psychosocial Rehabilitation Laboratory, Center for Rehabilitation Research (LabRP-CIR), Escola Superior de Saúde, Instituto Politécnico do Porto, 4200-072 Porto, Portugal; mpe@ess.ipp.pt (M.P.); ajmarques@ess.ipp.pt (A.M.)
2 Laboratory of Neuropsychophysiology, Faculty of Psychology and Education Sciences, University of Porto, 4200-135 Porto, Portugal; carinafernandes@fpce.up.pt (C.F.); fbarbosa@fpce.up.pt (F.B.)
3 Faculty of Human and Social Sciences, University Fernando Pessoa, 4249-004 Porto, Portugal
4 Molecular Oncology and Viral Pathology Group, Research Center of IPO (CI-IPOP) & RISE@CI-IPOP (Health Research Network), Portuguese Oncology Institute of Porto (IPO Porto)/Porto Comprehensive Cancer Center (Porto CCC), 4200-072 Porto, Portugal
* Correspondence: artemisa@ess.ipp.pt

Abstract: Background: Problematic social media (SM) use is a growing concern, particularly among adolescents who are drawn to these platforms for social interactions important to their age group. SM dependence is characterized by excessive, uncontrolled usage that impairs personal, social, and professional aspects. Despite the ongoing debate over recognizing SM addiction as a distinct diagnostic category, the impact of social feedback, particularly through the "like" button, on brain activity remains under scrutiny. Objective: This systematic review aims to study the neural correlates of online social feedback, focusing on the effects of the "like" feedback on brain activity using fMRI and EEG. Methods: The review followed the recommendations of the Preferred Reporting Items for Systematic Reviews and Meta-Analysis Protocols (PRISMA). Results: The review included 11 studies with 504 participants, identifying key brain structures such as the amygdala, ventromedial prefrontal cortex (vmPFC), and ventral striatum involved in reward processing. Positive feedback ("likes") activates areas like the nucleus accumbens (NACC), vmPFC, and amygdala, with NACC correlating with increased SM use intensity. Negative feedback activates the ventrolateral prefrontal cortex (vlPFC) and left medial prefrontal cortex (mPFC). Behavioral data indicates that positive feedback influences subsequent social interactions. Conclusions: The review highlights disparities in the literature regarding the neural response to social feedback, emphasizing the need for further research to clarify the roles of sex, personality traits, and the person giving feedback. Overall, understanding the neurobiological underpinnings of SM engagement is essential for developing effective interventions to prevent or address the negative effects of excessive SM use.

Keywords: social media; like button; EEG; fMRI

check for updates

Academic Editor: Daniele Giansanti

Received: 18 October 2024
Revised: 12 December 2024
Accepted: 30 December 2024
Published: 6 January 2025

Citation: Dores, A.R.; Peixoto, M.; Fernandes, C.; Marques, A.; Barbosa, F. The Effects of Social Feedback Through the "Like" Feature on Brain Activity: A Systematic Review. *Healthcare* **2025**, *13*, 89. https://doi.org/10.3390/healthcare13010089

1. Introduction

Problematic use of social media (SM) platforms is an ever-growing concern that affects various populations, especially people from low-income countries [1], with moderate to low school achievement, low parental control [2], low self-esteem [3], feelings of loneliness, fear of negative evaluation, and behavioral inhibition [4]. Adolescents seem to be amongst the more vulnerable populations due to SMs' unique appeal to their age range. They are

drawn to these platforms for social interactions [5]. However, social feedback plays a critical role in shaping social behavior and well-being throughout life, from childhood to older adulthood. In early development, positive social feedback from peers and caregivers helps foster social skills, self-esteem, and a sense of belonging. As individuals age, social feedback continues to influence behavior, reinforcing social bonds and maintaining relationships. In older adults, social feedback becomes increasingly important due to the reduction in social networks, leading to potential feelings of loneliness and isolation (e.g., Tragantzopoulou & Giannouli, [6]). Positive social interactions and feedback can mitigate these effects by promoting emotional well-being and reducing feelings of loneliness, while the absence or negative social feedback may exacerbate loneliness and contribute to mental health decline in this population.

SM dependence is defined as the excessive and uncontrolled usage of SMs, leading to impairments in personal, social, and professional aspects [7,8]. Excessive users of SM can show abstinence symptoms, such as anxiety [9,10], stress, and depression [9,11]. They can also develop a tolerance, resulting in the need to progressively increase the time spent on SMs to obtain the same level of gratification experienced in the early stages. Another relevant aspect of dependence is the excessive worry associated with SM-related behaviors, and a lack of self-control with internet usage [7]. However, other results show no negative outcomes, only showcasing some specific behaviors as potentially dangerous [12]. One of the reasons for this discrepancy is that SM dependence could also be explained based on the person's preferred SM, as some results show that changing behaviors could be more difficult in certain SM [13] Given these differences, this type of addiction is still being debated in the literature, with the Diagnostic and Statistical Manual of Mental Disorders (DSM-5; [14]) and International Classification Diseases (ICD-11; [15]) not recognizing addiction to SM as a separate diagnostic category, despite it being studied as a behavioral addiction when it becomes excessive and affects the daily life of an individual [8,16].

In the context of SM, the "like" button acts as a reward mechanism, providing quick and simple feedback on users' social media activity [17]. Given its purpose, the "like" feature can motivate users to adjust their publications and online behaviors to maximize the chances of receiving this reward [18].

The Interaction of Person-Affect-Cognition-Execution (I-PACE) model [19] offers insights into the underlying mechanisms involved in the development of various behavioral addictions. The I-PACE model highlights the importance of the interaction between predisposing factors, affective and cognitive responses to stimuli, and executive functions—such as inhibitory control and decision making—in the development of behavioral addiction. In the case of social media, like gambling or gaming disorders, the process may involve heightened cue-reactivity and craving, coupled with diminished inhibitory control, which can contribute to habitual behaviors. Further studies are needed to explore both the common and distinct mechanisms involved in addiction, obsessive-compulsive disorders, impulse control disorders, and substance use disorders, as these conditions share underlying neurobiological processes.

Studies have used functional magnetic resonance imaging (fMRI), electroencephalography (EEG), and evoked-related potentials (ERP) to better understand the effects of social feedback through the "like" on brain activity. Each technique offers advantages and limitations. fMRI has excellent spatial resolution, making it possible to identify the brain regions involved in the perception and processing of "likes" [20]. On the other hand, EEG, due to its high temporal resolution, is especially suitable for studying rapid cognitive processes such as attention and sensory responses.

fMRI studies show that the amygdala, ventromedial prefrontal cortex (vmPFC), and ventral striatum are key structures in reward processing [21–23]. The vmPFC plays a key

role in this network through the observation and evaluation of the reward (i.e., "number of likes"; followers), motivating behavior towards it. The amygdala further solidifies the behavior-oriented toward the positive outcome, which can lead to SM addiction. Additionally, using SM may be associated with reduced amygdala volume [22]. The striatum, which shares various connections with different brain structures, allows the integration of information from various modalities and is necessary for learning and evaluating rewards [21,23], which in turn supports motivated behavior [23]. The involvement of the striatum in reward processing extends to social situations, where certain situations, such as a compliment from a colleague, result in increased activity [21].

For EEG data, evoked potentials, specifically the P300 and N200 components, have been used to investigate online social interactions. The P300 is associated with the allocation of attention and processing of relevant stimuli [24–26]. In the context of social networks, an increase in the amplitude of the P300 suggests a greater allocation of neuronal resources to the processing of these socially relevant stimuli. The N200 is often associated with conflict detection and decision making [27].

Current research shows the activation of several structures related to reward processing and the impact that their activation has on continued engagement with SM. As such, the need for research that further validates the effects of certain aspects, such as the "like" feature, on continued engagement is increasingly present. Despite some online social interactions having similarities with offline interactions, they are not identical. For instance, reward processing differs, as online social rewards lead to higher activation of structures such as the nucleus accumbens (NACC) while showing less amygdala recruitment [22]. Identifying these differences, we aim to systematically review studies on the neural correlates of online social feedback, using EEG and fMRI as investigative techniques. The main aim is to systematize the neuronal correlates and neurobiological processes that underlie the response to social feedback, focusing on how this stimulus is processed by the brain structures involved in social reward processing. Considering the relevance of investigating the "like" feature of social media platforms through fMRI and EEG (e.g., Meshi et al. [28]; Bhanji & Delgado [21]), the following are the proposed main research questions:

Research question 1: What are the effects of the "like" feedback on neuronal regions associated with reward processing, assessed through fMRI?

Research question 2: What are the effects of the "like" feedback on brain activity, assessed through EEG?

Additional questions involve exploring potential differences in brain activity associated with feedback valence (i.e., positive versus negative) and comparing habitual versus sporadic social media users. Behavioral studies on this topic will also be reviewed.

2. Method

The recommendations of Preferred Systematic Review and Meta-analysis (PRISMA; [29]) were followed to guide the general stages and protocols of this review. The PRISMA 2020 Checklist can be found in Appendix A.

2.1. Search Strategy

This review was performed according to the actualized Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [29]. Articles published until September 2024 were selected from PubMed, Web of Science, and EBSCOhost (including the Academic Search Complete, Psychology and Behavioral Sciences Collection, CINAHL Plus with Full Text, Fonte Acadêmica, MedicLatina, PsycARTICLES, PsycBOOKS, and PsycINFO databases).

The search expression was as follows: "("like feedback" OR "like symbol" OR "social media feedback" OR "social feedback") AND (fMRI OR "functional magnetic resonance imaging" OR EEG OR electroencephalography OR "Event-related Potentials" OR ERP) AND ("reward processing" OR "reward system" OR "brain regions" OR "neural regions" OR "ventral striatum" OR "nucleus accumbens" OR "dopaminergic pathways" OR "reward circuitry") NOT (marketing OR "fake news")". To prevent publication and source selection bias, an additional manual search was performed.

2.2. Study Selection

We included observational cross-sectional studies that investigated the neural processing of social media-related feedback by samples of adolescents and young adults (12–35 years old). After being included in reporting research on the topic of the review, articles were excluded according to the following criteria: (a) articles without a group of adolescents or young adults processing social media-related feedback (criterion 1: other population); (b) articles that did not investigate the neural processing of social media-related feedback (criterion 2: other measures); (c) inaccessible articles or articles without information about the neural processing of social media-related feedback (criterion 3: lack of data); (d) articles published in other languages than English (criterion 4: inaccessible language); and (e) reviews, commentaries, case series, or methods (criterion 5: wrong publication type). Articles reporting only duplicated data were also excluded, and when articles reported an expansion of previously conducted research, data from the most recent article were selected (criterion 6: duplicated data).

2.3. Screening and Selection of Records

The results of the literature search were compiled on Rayyan QCRI [30]. Following Higgins and Green's [31] guidelines, after the elimination of duplicates, two researchers blindly screened the titles and abstracts, excluded the articles that were out of topic, and retained the remaining studies. When this task was completed, the screening was unblinded. The reference list of the included empirical studies and reviews were also screened, retaining titles in the topic that did not appear in the systematic search. Two authors read all the retained studies and, independently, decided to include or exclude them. Disagreements in both stages were solved by consensus.

2.4. Quality Assessment

Two independent reviewers assessed the quality of the included studies using the Appraisal Tool for Cross-Sectional Studies (AXIS), a recent tool developed to assess the methodological quality of observational cross-sectional studies [32]. AXIS contains 20 questions regarding the introduction, methods, results, and discussion of each study. Each question could be answered with "yes" (1 point) or "no"/"don't know" (0 point). Disagreements were solved by consensus.

2.5. Data Collection and Analysis

The data of each included article were added to an extraction sheet developed for this review and refined when necessary. When available, the following variables were extracted from each article: year of publication; sample size (including males and females); mean age and standard deviation; neural imaging method; goals of the study, self-report measures used; details of the experimental task; and neurophysiological and behavioral results.

3. Results

3.1. Search Results

A total of 203 articles were identified through the search string. Initially, 158 articles were excluded as duplicates. A total of 50 articles were then analyzed based on their titles and abstracts. Of these, 38 articles were excluded due to the following criteria: (a) out of topic (n = 32); (b) other age range (n = 2); (c) lack neuroimaging techniques (n = 2); (d) wrong publication type (n = 2). Twelve articles were selected for full-text analysis, of which five studies were included. An additional 14 articles were hand-searched, of which seven were excluded due to the following criteria: (a) systematic review (n = 4); (b) not relevant to the topic (n = 1); (c) outside of the age range (n = 1); and (d) not using neuroimaging techniques (n = 1). Seven articles were selected for a full-text analysis from this batch, of which six studies were included. In total, 12 articles were included in the final review. Cohen's kappa was used to assess inter-rater agreement, with a score of 0.93 indicating substantial agreement [33]. The study selection process is illustrated in the PRISMA flow diagram (Figure 1).

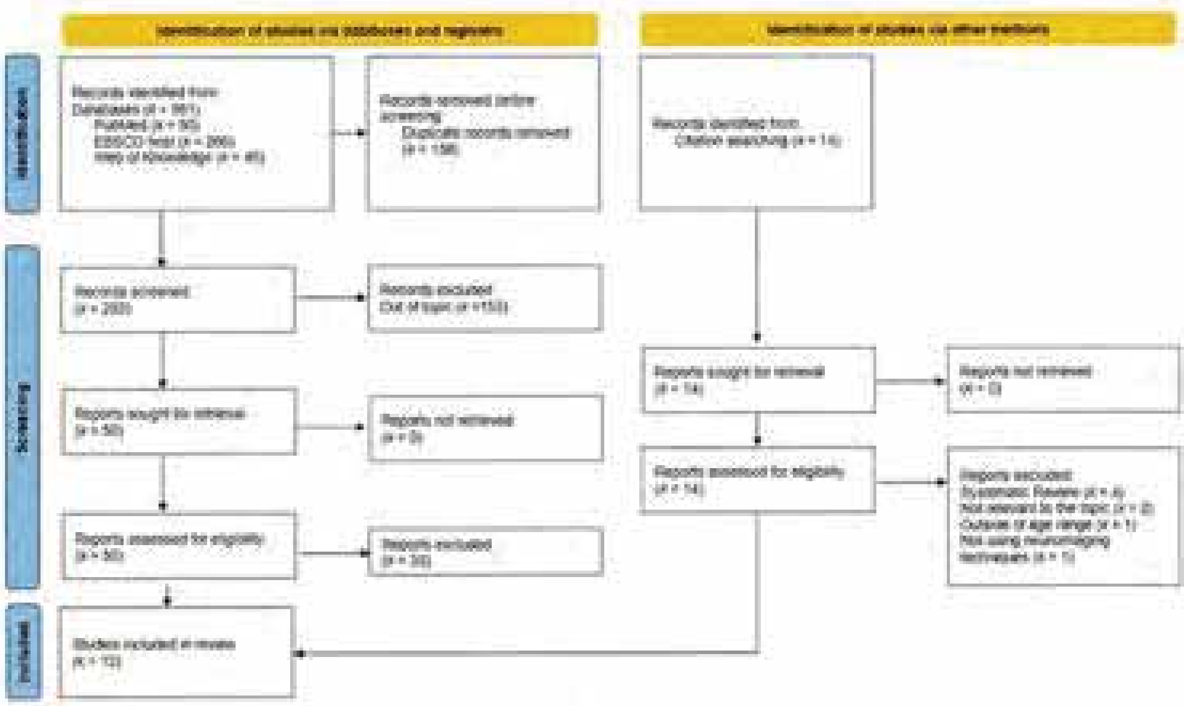


Figure 1. PRISMA flow diagram.

3.2. Studies Characteristics

The reviewed studies included a total of 537 participants, with 195 males and 295 females. One study did not report the participants' sex [34]. The techniques used to assess the participants were fMRI (n = 9); EEG (n = 3); and PET scan (n = 1). Regarding questionnaires, four studies [17,34–36] did not report using any questionnaire. The remaining studies [25,28,37–41] used a variety of instruments to assess their variables, with no questionnaire being consistently used across the literature. The methods used in the reviewed studies are presented in Table 1.

Table 1. Summary of the methodology used to assess online feedback.

Authors	Methods						
	Participants (N)	Gender (F/M)	Age (M ± SD) Min—Max	Neural Imaging Methods	Study Objectives	Self-Report Measures	Experimental Task
Somerville et al. [34]	42	Not reported	Not reported	fMRI	Testing social rejection and expectancy violation.	Not reported	Participants were initially photographed to have their picture rated by another; the rating was manipulated. They would view unfamiliar faces and had to say (yes or no) if they would like them based on first impressions. A second question was asked if the person in the photo would like the participant (yes or no). The same faces would appear after some time, with some having the opposite of the feedback initially given to them by the participant.
Izuma et al. [39]	19	F = 10 M = 9	21.6 ± 1.5	fMRI Pet scan	Testing if a good reputation activates reward-based circuits, and if these circuits are the same as those for monetary rewards.	Social Desirability Scale; Impressions Management Scale; Rosenberg Self-esteem Scale (RSES); NEO Five-Factor Inventory.	Participants would talk about a prevalent topic, talk about themselves, and take pictures. This material would then be rated by another group. In the second phase, predetermined feedback (positive or negative) was given.
Davey et al. [35]	19	F = 12 M = 7	19 ± 2.9 15–24	fMRI	Testing the activation of reward-related regions to receive a "like". Testing the effects of positive feedback and being "liked" by the opposite gender on certain brain regions.	Not reported	Participants had their pictures taken to be rated by another; the rating was manipulated. The participants thought they would view a set of neutral faces of other participants, which were from a pre-existing database. They would classify these faces based on the likelihood of liking the other person, and afterward, the faces appeared associated with a rating, supposedly given to the participant by the person in the photo.
Gunther et al. [37]	Total: 57 Prepubertal children: 12 Early adolescents: 14 Older adolescents: 15 Young adults: 16	Prepubertal children: F = 7 M = 5 Early adolescents: F = 8 M = 6 Older adolescents: F = 7 M = 8 Young adults: F = 8 M = 8	Prepubertal children: 9.7 ± 0.9 8–10 Early adolescents: 13.3 ± 0.8 12–14 Older adolescents: 17.1 ± 0.6 16–17 Young adults: 21.7 ± 1.9 19–25	fMRI	Testing the neural correlates of social acceptance and rejection in relation to age.	Raven's Progressive Matrices; Resistance to Peer Influence Questionnaire; Multidimensional Anxiety Scale for Children; Self-perception Profile for Children; Self-perception Profile for Adolescents.	Participants viewed unfamiliar age-matched faces and had to say (yes or no) if the person in the photo would like them based on first impressions. The same faces would appear after some time, with half having the opposite of the feedback initially given to them by the participant.

Table 1. *Cont.*

Authors	Methods						
	Participants (*N*)	Gender (F/M)	Age (*M* ± *SD*) Min—Max	Neural Imaging Methods	Study Objectives	Self-Report Measures	Experimental Task
Meshi et al., 2013 [28]	31	F = 17 M = 14	23.1 ± 3.2	fMRI	Testing if differences in the activation of NACC between social gains for the self and for others predicts SM use.	Facebook Intensity Scale; Structured Interview; RSES; Reynolds Social Desirability Scale-C; Narcissistic Personality Inventory; Mehrabian Conformity Scale; Beck Depression Inventory-II.	Participants saw their picture and a photo of another person and had to correctly identify whose picture it was. After answering, a word would appear that would correspond to a third-party assessment of the picture, this was manipulated. The feedback given was either positive or neutral.
Oumeziane et al. [42]	33	F = 19 M = 14	M = 26.34, SD = 6.25	EEG/ERP	To compare the neural processing of social (likes) and monetary rewards	Not reported	In the monetary task, participants anticipate and react to monetary rewards or punishments by responding quickly to specific cues, typically involving a delay period between stimulus and response. The social version of the task is similar, but the participants received likes instead of money.
Sherman [36]	32	F = 18 M = 14	13–18	fMRI	Testing the influence of "likes" and photos depicting risky behaviors on areas responsible for cognitive control.	Not reported	Initial submission of Instagram photos by participants. While in the fMRI, they viewed neutral images, images with risky and non-risky behaviors, and images submitted by other participants. The photos were displayed with either a high or low number of "likes". Half of the participants' photos had a high number of "likes", while the other half had few "likes". In the second version, the photos with a high number of "likes" had a lower number, and vice versa. When viewing the photos, the participants had to either "like" or skip them.
Sherman, Hernandez et al. [17]	58	F = 34 M = 24	18.2 13–21	fMRI	Testing the neural correlates of giving and receiving a "like".	Not reported	Initial submission of Instagram photos by participants. The participants viewed neutral images, images with risky behaviors, and images submitted by the participants, and had to either "like" or skip them.

Table 1. *Cont.*

Authors	Methods						
	Participants (N)	Gender (F/M)	Age ($M \pm SD$) Min—Max	Neural Imaging Methods	Study Objectives	Self-Report Measures	Experimental Task
Sherman, Greenfield et al. [40]	Total: 58 High schoolers: 32 University students: 26	F = 35 M = 23	High schoolers: 16.8 ± 1.4 University students: 19.9 ± 1.1 Overall 13–21	fMRI	Testing the neural effects of receiving many "likes". Testing age-related differences in the neural activation of reward processing and executive function.	Revised Cognitive Appraisal of Risky Events.	Initial submission of Instagram photos by participants. The participants viewed the "likes" on their pictures given by others, with "likes" being manipulated. While in the fMRI, they viewed neutral images, images with risky behaviors, and images submitted by participants. Half of the participants' photos had a high number of "likes", while the other half had few "likes". In the second version, the photos with a high number of "likes" had a lower number, and vice versa.
Nasn et al. [25]	77	F = 59 M = 18	20.8 ± 3.73	EEG	Testing the correlation between narcissistic traits and the fulfillment of narcissistic needs with social network engagement.	40-item Narcissistic Personality Inventory.	Participants used their phones with the Instagram app. The participants experienced social exclusion by performing a Cyberball task. The participants then performed a startle task, after which they were assigned to one of three conditions. For the selfies with likes condition, the participants posted a selfie with popular hashtags and were able to see the "likes" obtained in real-time; this was manipulated. For the selfie-only condition, the participants uploaded a selfie, and then viewed the picture for 5 min with no feedback. The neutral picture condition only stared at a picture of gravel.
Issa and Jabbouri [38]	19	F = 14 M = 5	19–23	EEG ECG	Testing neurophysiological responses to "likes".	Sociodemographic Questionnaire.	Participants were divided into three conditions: frequently receiving "likes"; less frequently "likes"; and receiving a moderate number of "likes". The participants sent 30 recently posted pictures of themselves. These pictures were mixed with pictures from other participants. They would then have to "like" or "skip". Throughout the experiment, several controlled notifications were sent with the number of "likes" received and the participants' positions on a scale.

Table 1. *Cont.*

Authors	Methods						
	Participants (N)	Gender (F/M)	Age (M ± SD) Min—Max	Neural Imaging Methods	Study Objectives	Self-Report Measures	Experimental Task
Wikman et al. [41]	Total: 92 Sample 1: 26 Sample 2: 66	Sample 1: F = 10 M = 16 Sample 2: F = 42 M = 24	Total: 18.7 ± 0.72 17–20	fMRI	Replicating previous findings on neural activation with virtual social interactions. Testing the relation between brain activity towards peer feedback and SM use.	Open-ended questionnaire about social network use.	Participants posted opinions on a Facebook group, receiving feedback from other participants. They were told that the feedback was prerecorded authentic feedback to the posted opinions. The participants were exposed to several controversial opinions and based on their answer (agree or disagree) a post in line with their answer was shared. The response given by peers, was either positive or negative, with some neutral reactions. For the control condition a neutral, factually correct statement was presented; the participant would then say true or false. The responses to the comment would also be neutral. The statement would appear for 3 s, followed by a 3 s window to answer, and by the post for 3 s.

Notes: Social media platform (SM); female (F); male (M); functional magnetic resonance imaging (fMRI); Electroencephalogram (EEG); Electrocardiogram (ECG); Electrooculogram (EOG).

The assessment of the studies using the AXIS (Appendix A) revealed that scores varied between 14 and 18 ($M = 15.8$; $SD = 1.17$). Several common weaknesses were observed across the majority of the studies. A significant issue was the lack of sample size justification, as most studies did not provide a clear rationale for their chosen sample sizes. This limitation impacts the statistical power of their conclusions and raises concerns about the reliability of their results. Additionally, power calculations were not explicitly reported in any of the studies, further weakening the robustness of the findings. Another recurring problem was the failure to identify potential biases. Few studies adequately discussed or accounted for biases that might influence their results, which is crucial for ensuring the accurate interpretation of the data. Furthermore, while most studies demonstrated strengths in terms of replicability, only a few addressed the generalizability of their results beyond the specific populations they studied, limiting the broader applicability of their findings. Despite these weaknesses, the studies generally excelled in terms of the clarity of their objectives, the appropriateness of their study designs, and the adequate presentation of results, reflecting the overall high standards in these areas.

3.3. *Neurophysiological and Behavioral Results*

Regarding the neurophysiological results (Table 2), several studies investigated the effect of social rewards/rejection with the SM context through both positive (i.e., "like") and negative feedback (i.e., negative comment, "dislike"). The structures sensitive to social feedback, regardless of valence, are the striatum [28,39], thalamus, cerebellum [39], ventral anterior cingulate cortex (ACC) [34], ventrolateral prefrontal cortex (vlPFC), medial prefrontal cortex (mPFC), occipital cortex, and superior temporal gyrus [41]. Specifically, for receiving a "like", the striatum, thalamus, ventral tegmental area (VTA), mPFC, motor cortex, occipital cortex, and cerebellum are activated [17].

Table 2. Summary of the neurophysiological and behavioral results.

Authors	Methodology		Results	
	Participants (N)	Technique	Neurophysiological	Behavioral
Somerville et al. [34]	42	fMRI	Dorsal ACC: sensitive to expectation violation Ventral ACC: sensitive to feedback valence (positive/negative)	-
Izuma et al. [39]	19	fMRI	Striatum, thalamus, and cerebellum: activation for both social and monetary rewards Caudate nucleus and putamen: similar activation for monetary and social rewards	Trials with positive feedback were more desirable.
Davey et al. [35]	19	fMRI	NACC, ventral midbrain, vmPFC, mid-cingulate cortex, amygdala, dorsal, ventral, and retrosplenial PCC: activation for positive feedback Amygdala, and vmPFC: higher activation for positive feedback received from a person with a higher rating Right OFC and right anterior insula: sensitive to gender	Female faces were rated as more positive than male faces, and faces rated higher were more rewarding to view.
Gunther et al. [37]	57	fMRI	Ventral mPFC, striatum, left subcallosal cortex, ACC, MCC, right putamen, right amygdala, left hippocampus, and right parahippocampal gyrus: activation for expectation violation, in adults Ventral mPFC, right subcallosal cortex, right PCC, left OFC, caudate nucleus, left precuneus, and left thalamus: higher activation for expected positive feedback compared to unexpected negative feedback Right subcallosal cortex, left caudate, right putamen, right middle frontal gyrus, and right inferior frontal gyrus: higher activation for expected negative feedback compared to unexpected positive feedback	Young adults chose "yes" more often than prepubertal children and early adolescents and were faster to respond than prepubertal children.
Meshi et al., 2013 [28]	31	fMRI	Ventral striatum: activation for both social and monetary rewards Left NACC: activation to positive feedback, correlated with Facebook usage intensity	-
Sherman [36]	32	fMRI	Visual cortex, and cerebellum: higher activation for neutral photos with many "likes" Left frontal cortex, precentral, middle frontal, and inferior frontal gyrus: higher activation for risky photos with many "likes" Precuneus, mPFC, lateral occipital cortex, hippocampus, NACC, caudate, putamen, thalamus, ventral tegmental area, and brain stem: higher activation for self-photos with many "likes" NACC: higher activation for photos with high amounts of "likes", except risky photos	Participants were more likely to "like" photos liked by others, and to not "like" unpopular photos.
Oumeziane et al. [42]	33	EEG/ERPs	Social and monetary rewards elicit comparable ERP latencies and scalp topographies across several processing stages (reward cue, outcome anticipation, and outcome evaluation), highlighting the possibility of a common neural network.	-
Sherman, Hernandez et al. [17]	58	fMRI	Striatum, thalamus, VTA, mPFC, motor cortex, occipital cortex, and cerebellum: activation for receiving many "likes"	Neutral images were more likely to be "liked" than pictures depicting faces. Appeal, being funny, and being similar to a picture taken by the participant were the more frequent motives for liking a picture. A visceral reaction was a more common reason to give a "like".
Sherman, Greenfield et al. [40]	58	fMRI	NACC: activation for self-photos with many "likes"	Popular images received more "likes" than unpopular ones. This effect was more prominent for self-pictures.

Table 2. *Cont.*

Authors	Methodology		Results	
	Participants (*N*)	Technique	Neurophysiological	Behavioral
Nash et al. [25]	77	EEG	P3: decreased amplitude on a selfie with "likes" compared to only "selfie" and the neutral condition, in people with high narcissism and higher scores of leadership/authority; increased amplitude for a selfie with "likes" compared to the neutral condition, in narcissistic people with low scores of leadership/authority	-
Issa and Jabbouri [38]	19	EEG ECG	Beta wave: predominant when receiving many or few "likes" Alpha waves: predominant when receiving moderate/average number of "likes" ECG: similar activity when receiving many or few "likes"	-
Wikman et al. [41]	92	fMRI	vlPFC, mPFC, occipital cortex, and superior temporal gyrus: activation for positive and negative feedback vlPFC; anterior insula; and left mPFC: activation for negative feedback Left posterior insula, medial superior parietal lobe, precuneus, PCC, and superior frontal gyrus: activation for positive feedback	Female participants spent more time on social media and received more "likes" per photo.

Notes. Anterior cingulate cortex (ACC); nucleus accumbens (NACC); ventromedial prefrontal cortex (vmPFC); posterior cingulate cortex (PCC); orbitofrontal cortex (OFC); ventral tegmental area (VTA); medial prefrontal cortex (mPFC); social media platform (SM); middle cingulate cortex (MCC); ventrolateral prefrontal cortex (vlPFC); Appraisal tool for Cross-Sectional Studies (AXIS).

Some studies also focus on specific activation towards positive and negative feedback. For positive feedback, the results showed the activation of the left posterior insula, medial superior parietal lobe, precuneus [41], NACC, ventral midbrain, vmPFC, mid-cingulate cortex, amygdala [35], and dorsal, ventral, and retrosplenial posterior cingulate cortex (PCC) [35,41]. Furthermore, the NACC activation to positive feedback was positively correlated with SM use intensity [28]. For negative feedback, the results showed the activation of the vlPFC and left mPFC [41]. Feedback can also come in the form of many or few "likes", where many "likes" are shown to activate the mPFC, lateral occipital cortex, hippocampus, NACC, caudate, putamen, thalamus, and VTA [36]. In terms of EEG activity, receiving many or few "likes" showed a predominance of beta waves, while receiving a moderate amount of "likes" showed a predominance of alpha waves [38].

The person giving the feedback can also influence brain activity through social feedback. A person more positively viewed led to higher activity in the amygdala and vmPFC [35]. Sex also influences brain activity, specifically, receiving positive feedback from the opposite sex results in higher activation of the right orbitofrontal cortex (OFC) and right anterior insula [35]. Expectations also play a role in understanding the neurological activity elicited by feedback. Positive feedback consistent with expectations activates the ventral mPFC, right subcallosal cortex, right PCC, left OFC, caudate nucleus, left precuneus, and left thalamus in contrast to unexpected negative feedback. Negative feedback consistent with expectation activates the right subcallosal cortex, left caudate, right putamen, right middle frontal gyrus, and right inferior frontal gyrus in contrast to unexpected positive feedback. Age, resistance to peer pressure, and social anxiety levels are also shown to influence these results [37]. A narcissistic personality trait also influences brain activity, with lower P3 amplitudes when receiving many "likes" on a selfie [25].

The reported behavioral results (Table 2) corroborate neuroimaging data, showing that positive feedback was more desirable [39]. When it comes to giving a "like", sex can influence the motivation to do so, as female faces are rated as more positive, which in turn makes them more rewarding to view, because higher-rated faces are reported to be more

rewarding [35]. This is consistent with the results showing that women receive more "likes" per photo [41].

4. Discussion

This systematic review aimed to analyze neurophysiological studies (i.e., studies using fMRI and EEG methodologies) investigating online social reward processing, mainly the feedback provided through the "like" feature.

The review found significant variability in the objectives, methods, and results across the studies. The objectives often align with key social factors influencing social media (SM) use, such as social rewards [13,43], usually in the form of a "like", reward learning, and decision making [23].

Regarding the main research questions—what are the effects of the "like" feedback on neuronal regions associated with reward processing, assessed through fMRI and through EEG?—the analysis of the data related to receiving feedback, both positive and negative, revealed activation in several brain structures associated with reward processing, including the striatum and thalamus [28,39]. The structures involved in goal-oriented and social behavior, such as the ventrolateral prefrontal cortex (vlPFC) and medial prefrontal cortex (mPFC) [44,45], are also engaged in feedback processing. These findings align with the existing literature that underscores the role of these structures in social reward processing [21,22,46,47].

Differential brain activity for positive versus negative feedback was also identified. The nucleus accumbens (NACC) appears particularly crucial for engagement with SM, as its activation correlates with positive feedback and the intensity of SM use [28]. This supports the notion that the NACC plays a significant role in motivating behavior related to positive social gains and avoiding social punishment [48]. The insula is involved in reward anticipation and avoiding social punishment [46], a finding corroborated by the reviewed data [41].

The amygdala's activation in response to positive feedback is expected [33], given its role in social behavior coordination [49] and positive outcome processing [22]. In contrast, the results for negative feedback were less extensive, primarily showing activation in the vlPFC and mPFC [41]. This gap highlights the need for future research into the neuroanatomical areas involved in processing negative feedback. Broader research on social rewards has identified the NACC, insula, and right inferior frontal gyrus as sensitive to negative social outcomes [46,48,50].

Specifically, when it comes to receiving a "like", the striatum, thalamus, hippocampus, and VTA are activated [17]. This is consistent with their roles in various aspects of social rewards and behavior [21,23,46,51]. Notably, the insula, amygdala, ventromedial prefrontal cortex (vmPFC), and paracingulate cortex are uniquely activated in response to receiving a "like" rather than giving feedback [18]. However, it is worth noting that some expected structures, such as the NACC, which is typically involved in processing positive feedback [28], were not identified in this study [17].

The significance of the "like" feature as social feedback is further supported by studies examining the effects of the quantity of "likes" on brain activity. Similar activation patterns were observed in structures such as the mPFC, NACC, and thalamus [36]. Additional structures activated by receiving many "likes" include the caudate, brain stem, VTA, and hippocampus [36]. The VTA and hippocampus are associated with memory and prosocial behavior [47,51,52] and contextual memory for social rewards [53], respectively. The caudate and brain stem are also integral to models of social reward processing [46], with the caudate being involved in both monetary and social rewards [54]. Despite their roles in social behavior, these regions were not consistently reported, with some appearing only

in specific paradigms related to feedback valence or quantity of positive feedback (i.e., number of "likes"). Understanding these discrepancies is crucial for elucidating the impact of social feedback on SM.

Furthermore, EEG data revealed predominantly beta wave activity in response to receiving many "likes" [38]. Given beta waves' involvement in decision making and selective attention [55], EEG findings corroborate fMRI data, highlighting the activating effect of numerous "likes".

Overall, the results show considerable variability in effects, and while not contradictory, they do not establish a consistent activation system for online social rewards. Methodological differences and various influencing factors, such as the characteristics of the feedback giver and recipient, including sex and personality traits, can affect reactions to positive feedback. For example, a more positive rating from the feedback giver was associated with the amygdala and vmPFC [35]. This suggests that higher social hierarchy may involve greater amygdala activation due to its role in hierarchical learning and reward processing [56]. The vmPFC's role in processing social feedback and self-relevant information [57] might explain its activation when receiving feedback from individuals with higher ratings. Additional structures, such as the hippocampus and mPFC, also relate to social hierarchy [56], indicating a need for further research into their involvement in feedback processing from high-status individuals. While evidence on sex differences in processing social feedback is limited, one study indicated increased activity in the right OFC and right anterior insula in response to feedback from the opposite sex [35]. The particular activation of these structures could be related to somatic activation [58] or sexually relevant stimuli, such as attraction toward the opposite sex [59]. Personality traits may also influence SM engagement through rewards tailored to specific traits, affecting brain activity such as the P3 component [25].

Non-neuroimaging studies reinforce the importance of the characteristics of the giver and recipient of the "like". On one hand, the characteristics of the giver are shown to be more relevant in determining the subjective value of the "like" than the number of "likes" received [60]. On the other, the characteristics of the receiver (i.e., age) seem to alter the perceived value of the "like" [61]. Personality is one of these characteristics, with certain personality profiles, such as narcissism [62], leading to higher SM use [63,64], further modulated by sex and age [65]. Indeed, these individual characteristics seem to be potential vulnerabilities towards a pathological engagement with SM, as several studies have shown that variables such as impulsivity, narcissism, fear of missing out, and low self-esteem can be associated with an increased risk of developing this kind of involvement [66]. As such, and as pointed out in some studies [35] the neural correlates of social feedback could vary based on the characteristics of the giver and recipient.

The findings from our systematic review offer valuable insights into the neural correlates of social media feedback, particularly the "like" feature, and its impact on brain activity.

One of the key clinical insights is the role of brain structures such as the nucleus accumbens (NACC), vmPFC, and amygdala in reward processing. The activation of these regions in response to positive social feedback highlights the neurobiological underpinnings of social media use, which can inform clinical approaches to addressing problematic social media behavior, particularly in adolescents. The excessive activation of these reward-related brain areas, especially the NACC, is linked to increased intensity of social media use, which may contribute to the development of dependency or addiction-like behaviors [28]. This suggests that interventions aimed at reducing social media use could benefit from targeting the reward systems and enhancing self-regulation strategies to counterbalance excessive reward-seeking behaviors.

The review highlights the amygdala's role in social behavior and its involvement in processing positive feedback. Given that the amygdala is also associated with emotional regulation, its activation in the context of social media may explain why some individuals are more prone to anxiety or mood disturbances in response to online interactions [22]. Clinically, this underscores the need for mental health interventions that address the emotional impact of online social interactions, particularly in populations vulnerable to anxiety or depression.

Additionally, the findings suggest a significant difference in how positive and negative feedback is processed in the brain. The relatively limited involvement of brain structures in response to negative feedback, particularly when compared to the robust activation seen with positive feedback, may indicate that users of social media, particularly adolescents, are more sensitive to social rewards than to social punishments. Clinically, this raises concerns about how the continuous pursuit of social validation may reinforce maladaptive behaviors, such as excessive engagement with social media, while potentially diminishing resilience to social rejection or negative feedback [41].

There are similarities between the brain structures involved in processing the "like" feedback and the structures involved in processing gambling outcomes. For example, the activation of the amygdala and the ventromedial prefrontal cortex (vmPFC) to gambling outcomes could be an indicator of problematic gambling [67,68]. Non-neuroimaging data corroborates the similarities between conditions, as evidenced by increased decay of the reward value over time (delay discounting) [69,70], or by problematic engagement possibly serving as a coping mechanism with adverse consequences [19,71]. Furthermore, individuals with more severe SM-related problematic behavior are more likely to exhibit additional problematic behaviors or even develop a behavioral addiction, such as gambling disorder [72]. Despite the similarities, a "like" functions as a social reward, in contrast to the monetary reward associated with gambling outcomes. One key difference lies in the value of a "like", which is influenced by the characteristics of the giver. However, to our knowledge, no studies have directly compared these two types of rewards to clarify their differences. In light of these findings, future research should explore the longitudinal effects of social media engagement on mental health, especially in individuals with pre-existing vulnerabilities such as low self-esteem or social anxiety. Understanding how these neural patterns evolve could inform prevention strategies for adolescents and young adults at risk of developing dependency on social media platforms.

Clinically, the evidence from our review calls for the development of therapeutic interventions aimed at moderating the neural reward systems activated by social feedback. Cognitive–behavioral interventions focusing on self-regulation, emotional resilience, and the management of online interactions could help mitigate the potential negative impacts on mental health. Additionally, educational programs for adolescents, focusing on the psychological mechanisms behind social media usage, may empower them to navigate online spaces more consciously and healthily.

Despite the contribution of this study to the literature, some limitations must be considered when interpreting the results. One of the main limitations of this systematic review is the heterogeneity of methodologies across the included studies. The included studies used different neuroimaging techniques, including fMRI and EEG, each of which has distinct strengths and limitations. For instance, fMRI provides excellent spatial resolution, allowing the identification of specific brain regions involved in reward processing, but it lacks the temporal resolution needed to capture the fast dynamics of brain activity. On the other hand, EEG offers high temporal resolution but lacks spatial specificity, making it difficult to identify the exact brain regions responsible for certain neural responses. This variation

in neuroimaging methods hinders the direct comparison of findings across studies and introduces potential biases related to the limitations inherent to each technique.

Additionally, there was significant variability in the participant demographics across the studies, particularly regarding age, sex, and social media usage habits. Most of the studies included adolescents and young adults, but the age ranges varied, and some studies did not report detailed demographic data, such as sex distribution. This heterogeneity can influence the results, as neural responses to social feedback may differ by demographic differences, as well as personality traits or susceptibility to social influence. As a result, it is difficult to generalize the findings to a broader population or to make definitive conclusions about how different groups may respond to social media feedback. Also, the studies had relatively small sample sizes, which may affect the statistical power of their findings. These limitations make it difficult to synthesize the results and draw comprehensive conclusions. Overcoming these limitations in future research will be crucial for gaining a more complete understanding of how social media feedback influences brain activity.

5. Conclusions

The current findings on the neural correlates of online social feedback, focusing on the effects of the "like" feedback on brain activity, using EEG and fMRI, show a fragmented picture, with various findings that, while not contradictory, do not clarify a distinct network of brain structures involved in processing feedback, such as the "like" button. To better understand this, it is crucial to consider a wider range of influencing variables, including sex and personality traits, which have been shown to affect brain activity. Additionally, peripheral physiological indicators could offer valuable insights into users' reactions to SM and their subsequent behaviors.

It is important to note that giving or receiving a "like" involves different brain processes, and our review focused solely on the experience of receiving a "like". This limitation means that our findings might not fully capture the nuances observed in studies examining both giving and receiving "likes". Another limitation arises from some studies using alternative forms of positive feedback instead of the standard "like" button, which makes difficult the systematic analysis of results.

To enhance the validity of current findings, the increased use of EEG, due to its superior temporal resolution, would be beneficial. Comparing excessive SM users with regular or healthy users could also provide more precise insights into group differences.

In summary, as SM becomes increasingly integrated into daily life, more research is needed to explore the individual and SM characteristics, namely the role of the "like" feedback that drives user engagement. It is crucial to examine a wide range of variables to ensure that research remains relevant in the rapidly evolving field of social media, and the "like" feedback should be among them. Robust evidence is necessary to support and refine the theoretical models of online social rewards and to develop effective prevention and intervention strategies.

Author Contributions: A.R.D. and C.F. were responsible for the conceptualization of the research and development of the methodology. A.R.D. and M.P. contributed to the design and execution of the methods and conducted the formal analysis of the data. A.R.D. led the investigation and wrote with M.P. the original draft of the manuscript. F.B., C.F. and A.M. were involved in the review and editing of the manuscript. F.B. provided supervision throughout the research project. All authors have read and agreed to the published version of the manuscript.

Funding: CIR was supported by Fundação para a Ciência e Tecnologia [Portuguese Foundation for Science and Technology] (FCT) through R&D Units funding (UIDB/05210/2020). Miguel Peixoto was supported by Fundação para a Ciência e Tecnologia [Portuguese Foundation for Science and Technology] (FCT) (PRT/BD/154830/2023).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors have no conflicts of interest to disclose.

Appendix A

Table A1. Appraisal Tool for Cross-Sectional Studies (AXIS).

	Study											
AXIS	1	2	3	4	5	6	7	8	9	10	11	12
Introduction												
1	1	1	1	1	1	1	1	1	1	1	1	1
Methods												
2	1	1	1	1	1	1	1	1	1	1	1	1
3	0	0	0	0	0	0	0	0	0	0	0	1
4	1	1	1	1	1	1	1	1	1	1	1	1
5	0	0	0	0	0	0	0	0	0	0	0	1
6	0	0	0	0	0	0	0	0	0	0	0	0
7	1	1	1	1	1	1	1	1	1	1	1	1
8	0	0	0	0	0	0	0	0	0	0	0	0
9	1	1	1	1	1	1	1	1	1	1	1	1
10	1	1	1	1	1	1	1	1	1	1	1	1
11	1	1	1	1	1	1	1	1	1	1	1	1
Results												
12	1	1	0	0	1	1	1	1	1	1	0	1
13	1	1	1	1	1	1	1	1	1	1	1	1
14	1	1	1	1	1	1	1	1	1	1	1	1
15	1	1	1	1	1	1	1	1	1	1	1	1
16	1	1	1	1	1	1	1	1	1	1	1	1
Discussion												
17	1	1	1	1	1	1	1	1	1	1	1	1
18	1	1	0	0	1	1	1	1	1	1	0	1
Other												
19	1	1	1	1	1	1	1	1	1	1	1	1
20	1	1	1	1	1	1	1	1	1	1	1	1
Total	16	16	14	14	16	16	16	16	16	16	14	18

References

1. Casale, S.; Akbari, M.; Seydavi, M.; Bocci Benucci, S.; Fioravanti, G. Has the prevalence of problematic social media use increased over the past seven years and since the start of the COVID-19 pandemic? A meta-analysis of the studies published since the development of the Bergen social media addiction scale. *Addict. Behav.* **2023**, *147*, 107838. [CrossRef] [PubMed]
2. Paakkari, L.; Tynjälä, J.; Lahti, H.; Ojala, K.; Lyyra, N. Problematic social media use and health among adolescents. *Int. J. Environ. Res. Public Health* **2021**, *18*, 1885. [CrossRef] [PubMed]
3. Hou, Y.; Xiong, D.; Jiang, T.; Song, L.; Wang, Q. Social media addiction: Its impact, mediation, and intervention. *Cyberpsychology J. Psychosoc. Res. Cyberspace* **2019**, *13*, 4. [CrossRef]
4. Savci, M.; Aysan, F. #Interpersonal competence, loneliness, fear of negative evaluation, and reward and punishment as predictors of social media addiction and their accuracy in classifying adolescent social media users and non-users. *ADDICTA Turk. J. Addict.* **2018**, *5*, 431–471.
5. Nesi, J. The impact of social media on youth mental health: Challenges and opportunities. *North Carol. Med. J.* **2020**, *81*, 116–121. [CrossRef] [PubMed]
6. Tragantzopoulou, P.; Giannouli, V. Social isolation and loneliness in old age: Exploring their role in mental and physical health. *Psychiatriki* **2021**, *32*, 59–66. [CrossRef] [PubMed]
7. Kuss, D.; Griffiths, M. Social networking sites and addiction: Ten lessons learned. *Int. J. Environ. Res. Public Health* **2017**, *14*, 311. [CrossRef]

8. Mitropoulou, E.M.; Karagianni, M.; Thomadakis, C. Social media addiction, self-compassion, and psychological well-Being: A structural equation model. *Alpha Psychiatry* **2022**, *23*, 298–304. [CrossRef]
9. Bashir, H.; Bhat, S.A. Effects of social media on mental health: A review. *Int. J. Indian Psychol.* **2017**, *4*, 125–131. [CrossRef]
10. O'Day, E.B.; Heimberg, R.G. Social media use, social anxiety, and loneliness: A systematic review. *Comput. Hum. Behav. Rep.* **2021**, *3*, 100070. [CrossRef]
11. Cunningham, S.; Hudson, C.C.; Harkness, K. Social media and depression symptoms: A meta-analysis. *Res. Child Adolesc. Psychopathol.* **2021**, *49*, 241–253. [CrossRef]
12. Berryman, C.; Ferguson, C.J.; Negy, C. Social media use and mental health among young adults. *Psychiatr. Q.* **2018**, *89*, 307–314. [CrossRef] [PubMed]
13. Meshi, D.; Turel, O.; Henley, D. Snapchat vs. Facebook: Differences in problematic use, behavior change attempts, and trait social reward preferences. *Addict. Behav. Rep.* **2020**, *12*, 100294. [CrossRef] [PubMed]
14. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, 5th ed.; American Psychiatric Association: Washington, DC, USA, 2013. [CrossRef]
15. World Health Organization. *International Classification of Diseases*; World Health Organization: Geneva, Switzerland, 2018.
16. Fineberg, N.; Demetrovics, Z.; Stein, D.; Ioannidis, K.; Potenza, M.; Grünblatt, E.; Brand, M.; Billieux, J.; Carmi, L.; King, D.; et al. Manifesto for a European research network into problematic usage of the internet. *Eur. Neuropsychopharmacol.* **2018**, *28*, 1232–1246. [CrossRef]
17. Sherman, L.E.; Hernandez, L.M.; Greenfield, P.M.; Dapretto, M. What the brain 'Likes': Neural correlates of providing feedback on social media. *Soc. Cogn. Affect. Neurosci.* **2018**, *13*, 699–707. [CrossRef] [PubMed]
18. Mauri, M.; Cipresso, P.; Balgera, A.; Villamira, M.; Riva, G. Why Is Facebook so successful? Psychophysiological measures describe a core flow state while using Facebook. *Cyberpsychology Behav. Soc. Netw.* **2011**, *14*, 723–731. [CrossRef] [PubMed]
19. Brand, M.; Wegmann, E.; Stark, R.; Müller, A.; Wölfling, K.; Robbins, T.W.; Potenza, M.N. The Interaction of Person-Affect-Cognition-Execution (I-PACE) model for addictive behaviors: Update, generalization to addictive behaviors beyond internet-use disorders, and specification of the process character of addictive behaviors. *Neurosci. Biobehav. Rev.* **2019**, *104*, 1–10. [CrossRef]
20. Poudel, G.R.; Jones, R.D. Multimodal neuroimaging with simultaneous fMRI and EEG. In *Handbook of Neuroengineering*; Thakor, N.V., Ed.; Springer Nature: Singapore, 2022. [CrossRef]
21. Bhanji, J.P.; Delgado, M.R. The social brain and reward: Social information processing in the human striatum. *WIREs Cogn. Sci.* **2014**, *5*, 61–73. [CrossRef] [PubMed]
22. Doheny, M.M.; Lighthall, N.R. Social cognitive neuroscience in the digital age. *Front. Hum. Neurosci.* **2023**, *17*, 1168788. [CrossRef] [PubMed]
23. Fareri, D.S.; Delgado, M.R. Social rewards and social networks in the human brain. *Neuroscientist* **2014**, *20*, 387–402. [CrossRef] [PubMed]
24. Moretta, T.; Buodo, G. Motivated attention to stimuli related to social networking sites: A cue-reactivity study. *J. Behav. Addict.* **2021**, *10*, 314–326. [CrossRef] [PubMed]
25. Nash, K.; Johansson, A.; Yogeeswaran, K. Social media approval reduces emotional arousal for people high in narcissism: Electrophysiological evidence. *Front. Hum. Neurosci.* **2019**, *13*, 292. [CrossRef] [PubMed]
26. Polich, J. Updating P300: An integrative theory of P3a and P3b. *Clin. Neurophysiol.* **2007**, *118*, 2128–2148. [CrossRef]
27. Ye, Y.; Jiang, F.; Zhang, W. The neural and psychological processes of peer-influenced online donation decision: An event-related potential study. *Front. Psychol.* **2022**, *13*, 899233. [CrossRef]
28. Meshi, D.; Morawetz, C.; Heekeren, H.R. Nucleus accumbens response to gains in reputation for the self relative to gains for others predicts social media use. *Front. Hum. Neurosci.* **2013**, *7*, 57063. [CrossRef] [PubMed]
29. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **2021**, *372*, n71. [CrossRef] [PubMed]
30. Ouzzani, M.; Hammady, H.; Fedorowicz, Z.; Elmagarmid, A. Rayyan—A web and mobile app for systematic reviews. *Syst. Rev.* **2016**, *5*, 210. [CrossRef] [PubMed]
31. Higgins, J.; Green, S. *Cochrane Handbook for Systematic Reviews of Interventions*; John Wiley & Sons: Hoboken, NJ, USA, 2019.
32. Downes, M.J.; Brennan, M.L.; Williams, H.C.; Dean, R.S. Development of a critical appraisal tool to assess the quality of cross-sectional studies (AXIS). *BMJ Open* **2016**, *6*, e011458. [CrossRef] [PubMed]
33. McHugh, M.L. Interrater reliability: The kappa statistic. *Biochem. Med.* **2012**, *22*, 276–282. [CrossRef]
34. Somerville, L.H.; Heatherton, T.F.; Kelley, W.M. Anterior cingulate cortex responds differentially to expectancy violation and social rejection. *Nat. Neurosci.* **2006**, *9*, 1007–1008. [CrossRef]
35. Davey, C.G.; Allen, N.B.; Harrison, B.J.; Dwyer, D.B.; Yücel, M. Being liked activates primary reward and midline self-related brain regions. *Hum. Brain Mapp.* **2010**, *31*, 660–668. [CrossRef] [PubMed]

36. Sherman, L.E.; Payton, A.A.; Hernandez, L.M.; Greenfield, P.M.; Dapretto, M. The power of the like in adolescence: Effects of peer influence on neural and behavioral responses to social media. *Psychol. Sci.* **2016**, *27*, 1027–1035. [CrossRef] [PubMed]
37. Gunther Moor, B.; Van Leijenhorst, L.; Rombouts, S.A.R.B.; Crone, E.A.; Van Der Molen, M.W. Do you like me? Neural correlates of social evaluation and developmental trajectories. *Soc. Neurosci.* **2010**, *5*, 461–482. [CrossRef]
38. Issa, H.; Jabbouri, R. The infamous "Like" feature: A neuro perspective. *Int. J. Technol. Hum. Interact.* **2022**, *18*, 1–22. [CrossRef]
39. Izuma, K.; Saito, D.N.; Sadato, N. Processing of social and monetary rewards in the human striatum. *Neuron* **2008**, *58*, 284–294. [CrossRef] [PubMed]
40. Sherman, L.E.; Greenfield, P.M.; Hernandez, L.M.; Dapretto, M. Peer Influence via Instagram: Effects on brain and behavior in adolescence and young adulthood. *Child Dev.* **2018**, *89*, 37–47. [CrossRef]
41. Wikman, P.; Moisala, M.; Ylinen, A.; Lindblom, J.; Leikas, S.; Salmela-Aro, K.; Lonka, K.; Güroğlu, B.; Alho, K. Brain responses to peer feedback in social media are modulated by valence in late adolescence. *Front. Behav. Neurosci.* **2022**, *16*, 790478. [CrossRef]
42. Oumeziane, B.A.; Schryer-Praga, J.; Foti, D. "Why don't they 'like' me more?": Comparing the time courses of social and monetary reward processing. *Neuropsychologia* **2017**, *107*, 48–59. [CrossRef]
43. Wadsley, M.; Covey, J.; Ihssen, N. The predictive utility of reward-based motives underlying excessive and problematic social networking site use. *Psychol. Rep.* **2022**, *125*, 2485–2516. [CrossRef] [PubMed]
44. Sturm, V.E.; Haase, C.M.; Levenson, R.W. Emotional dysfunction in psychopathology and neuropathology: Neural and Genetic Pathways. In *Genomics, Circuits, and Pathways in Clinical Neuropsychiatry*; Academic Press: Amsterdam, The Netherlands, 2016; pp. 345–364. [CrossRef]
45. Xu, P.; Chen, A.; Li, Y.; Xing, X.; Lu, H. Medial prefrontal cortex in neurological diseases. *Physiol. Genom.* **2019**, *51*, 432–442. [CrossRef]
46. Martins, D.; Rademacher, L.; Gabay, A.S.; Taylor, R.; Richey, J.A.; Smith, D.V.; Goerlich, K.S.; Nawijn, L.; Cremers, H.R.; Wilson, R.; et al. Mapping social reward and punishment processing in the human brain: A voxel-based meta-analysis of neuroimaging findings using the social incentive delay task. *Neurosci. Biobehav. Rev.* **2021**, *122*, 1–17. [CrossRef]
47. Meshi, D.; Tamir, D.I.; Heekeren, H.R. The emerging neuroscience of social media. *Trends Cogn. Sci.* **2015**, *19*, 771–782. [CrossRef]
48. Kohls, G.; Perino, M.T.; Taylor, J.M.; Madva, E.N.; Cayless, S.J.; Troiani, V.; Price, E.; Faja, S.; Herrington, J.D.; Schultz, R.T. The nucleus accumbens is involved in both the pursuit of social reward and the avoidance of social punishment. *Neuropsychologia* **2013**, *51*, 2062–2069. [CrossRef]
49. Raam, T.; Hong, W. Organization of neural circuits underlying social behavior: A consideration of the medial amygdala. *Curr. Opin. Neurobiol.* **2021**, *68*, 124–136. [CrossRef] [PubMed]
50. Zinchenko, O. Brain responses to social punishment: A meta-analysis. *Sci. Rep.* **2019**, *9*, 12800. [CrossRef] [PubMed]
51. Hung, L.W.; Neuner, S.; Polepalli, J.S.; Beier, K.T.; Wright, M.; Walsh, J.J.; Lewis, E.M.; Luo, L.; Deisseroth, K.; Dölen, G.; et al. Gating of social reward by oxytocin in the ventral tegmental area. *Science* **2017**, *357*, 1406–1411. [CrossRef]
52. Nikulina, E.M.; Johnston, C.E.; Wang, J.; Hammer, R.P. Neurotrophins in the ventral tegmental area: Role in social stress, mood disorders and drug abuse. *Neuroscience* **2014**, *282*, 122–138. [CrossRef]
53. Duarte, J.M.; Li, K.; Nguyen, R.; Ciocchi, S. Hippocampal contextualization of social rewards. *BioRxiv* **2023**. [CrossRef]
54. Wardle, M.C.; Fitzgerald, D.A.; Angstadt, M.; Sripada, C.S.; McCabe, K.; Luan Phan, K. The caudate signals bad reputation during trust decisions. *PLoS ONE* **2013**, *8*, e68884. [CrossRef]
55. Kropotov, J.D. *Functional Neuromarkers for Psychiatry: Applications for Diagnosis and Treatment*; Elsevier/ Academic Press: Amsterdam, The Netherlands, 2016.
56. Li, S.; Krueger, F.; Camilleri, J.A.; Eickhoff, S.B.; Qu, C. The neural signatures of social hierarchy-related learning and interaction: A coordinate- and connectivity-based meta-analysis. *NeuroImage* **2021**, *245*, 118731. [CrossRef]
57. Hiser, J.; Koenigs, M. The multifaceted role of the ventromedial prefrontal cortex in emotion, decision making, social cognition, and psychopathology. *Biol. Psychiatry* **2018**, *83*, 638–647. [CrossRef] [PubMed]
58. Rudebeck, P.H.; Rich, E.L. Orbitofrontal cortex. *Curr. Biol.* **2018**, *28*, R1083–R1088. [CrossRef]
59. Ishai, A. Sex, beauty and the orbitofrontal cortex. *Int. J. Psychophysiol.* **2007**, *63*, 181–185. [CrossRef] [PubMed]
60. Scissors, L.; Burke, M.; Wengrovitz, S. What's in a Like?: Attitudes and behaviors around receiving Likes on Facebook. In Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing, San Francisco, CA, USA, 27 February–2 March 2016; pp. 1501–1510. [CrossRef]
61. Rawlings, L. The Power of the "Like": A Quantitative Study on the Facebook Emoji as Social Support. 2023. Available online: https://digitalcommons.liberty.edu/doctoral/5022 (accessed on 15 September 2023).
62. Brailovskaia, J.; Ozimek, P.; Rohmann, E.; Bierhoff, H.-W. Vulnerable narcissism, fear of missing out (FoMO) and addictive social media use: A gender comparison from Germany. *Comput. Hum. Behav.* **2023**, *144*, 107725. [CrossRef]
63. Kircaburun, K.; Alhabash, S.; Tosuntaş, Ş.B.; Griffiths, M.D. Uses and gratifications of problematic social media use among university students: A simultaneous examination of the big five of personality traits, social media platforms, and social media use motives. *Int. J. Ment. Health Addict.* **2020**, *18*, 525–547. [CrossRef]

64. Sumaryanti, I.U.; Azizah, S.; Diantina, F.P.; Nawangsih, E. Personality and social media addiction among college students. In *Proceedings of the 2nd Social and Humaniora Research Symposium (SoRes 2019)*; Atlantis Press: Bandung, Indonesia, 2020. [CrossRef]
65. Correa, T.; Hinsley, A.W.; De Zúñiga, H.G. Who interacts on the Web?: The intersection of users' personality and social media use. *Comput. Hum. Behav.* **2010**, *26*, 247–253. [CrossRef]
66. Sheinov, V.P.; Dziavitsyn, A.S. Relations of social media addiction with impulsivity, narcissism and assertiveness in Belarusian and Ukrainian university students. *RUDN J. Psychol. Pedagog.* **2022**, *19*, 233–252. [CrossRef]
67. Genauck, A.; Quester, S.; Wüstenberg, T.; Mörsen, C.; Heinz, A.; Romanczuk-Seiferth, N. Reduced loss aversion in pathological gambling and alcohol dependence is associated with differential alterations in amygdala and prefrontal functioning. *Sci. Rep.* **2017**, *7*, 16306. [CrossRef] [PubMed]
68. Potenza, M.N.; Balodis, I.M.; Derevensky, J.; Grant, J.E.; Petry, N.M.; Verdejo-Garcia, A.; Yip, S.W. Gambling disorder. *Nat. Rev. Dis. Primers* **2019**, *5*, 51. [CrossRef]
69. Weinsztok, S.; Brassard, S.; Balodis, I.; Martin, L.E.; Amlung, M. Delay discounting in established and proposed behavioral addictions: A systematic review and meta-analysis. *Front. Behav. Neurosci.* **2021**, *15*, 786358. [CrossRef] [PubMed]
70. Turel, O.; He, Q.; Brevers, D.; Bechara, A. Delay discounting mediates the association between posterior insular cortex volume and social media addiction symptoms. *Cogn. Affect. Behav. Neurosci.* **2018**, *18*, 694–704. [CrossRef] [PubMed]
71. Burrow, A.L.; Rainone, N. How many likes did I get?: Purpose moderates links between positive social media feedback and self-esteem. *J. Exp. Soc. Psychol.* **2017**, *69*, 232–236. [CrossRef]
72. Tullett-Prado, D.; Stavropoulos, V.; Gomez, R.; Doley, J. Social media use and abuse: Different profiles of users and their associations with addictive behaviours. *Addict. Behav. Rep.* **2023**, *17*, 100479. [CrossRef] [PubMed]

Research Article

aps
ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Psychological Science
2016, Vol. 27(7) 1027–1035
© The Author(s) 2016
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0956797616645673
pss.sagepub.com
SAGE

The Power of the *Like* in Adolescence: Effects of Peer Influence on Neural and Behavioral Responses to Social Media

Like

Lauren E. Sherman[1,2,3], Ashley A. Payton[4], Leanna M. Hernandez[2,4], Patricia M. Greenfield[1,3], and Mirella Dapretto[2,5]

[1]Department of Psychology, University of California, Los Angeles; [2]Ahmanson-Lovelace Brain M
University of California, Los Angeles; [3]Children's Digital Media Center @ Los Angeles, University
Los Angeles, and California State University, Los Angeles; [4]Interdepartmental Neuroscience Prog
University of California, Los Angeles; and [5]Department of Psychiatry and Biobehavioral Scienc
University of California, Los Angeles

Abstract

We investigated a unique way in which adolescent peer influence occurs
functional MRI (fMRI) paradigm to simulate Instagram, a popular social phot
behavioral and neural responses to *likes*, a quantifiable form of social en
influence. Adolescents underwent fMRI while viewing photos ostensibly
likely to *like* photos depicted with many likes than photos with few likes; th
peer endorsement and held for both neutral photos and photos of risky beh
photos with many (compared with few) likes was associated with greater activity in neural regions implicated in reward processing, social cognition, imitation, and attention. Furthermore, when adolescents viewed risky photos (as opposed to neutral photos), activation in the cognitive-control network decreased. These findings highlight possible mechanisms underlying peer influence during adolescence.

Keywords

adolescent development, social cognition, social influences, risk taking, neuroimaging, open materials

Received 9/12/15; Revision accepted 3/31/16

Social media are immensely popular among adolescents: Nearly 90% of American teens report being active users, and young people have continually outpaced other age groups in adopting new media (Lenhart, 2015). Given this prevalence, it is unsurprising that parents, educators, and the popular press have expressed concerns about the effects of social media on social-skill development and interpersonal interactions. Frequently, these concerns manifest themselves in questions about the effect of social media on the developing brain. Nonetheless, few studies have examined neural mechanisms underlying any kind of social-media use (Choudhury & McKinney, 2013; Mills, 2014).

The neural correlates of social-media use are particularly important to understand in the context of adolescence, and not only because adolescents are enthusiastic users. Adolescence is especially important for social cognitive development; it is theorized to be a sensitive period during which young people are uniquely attuned to the complexities of interpersonal relationships (Baird, 2012; Blakemore & Mills, 2014). Subcortical regions functionally associated with emotion processing and reward undergo considerable changes and reorganization during puberty (Brenhouse & Andersen, 2011; Sisk & Foster, 2004). The dopaminergic system and related regions in the striatum are implicated in

Corresponding Author:
Lauren E. Sherman, Department of Psychology, UCLA, 1285 Franz Hall, Box 951563, Los Angeles, CA 90095-1563
E-mail: lsherman@ucla.edu

Research Article

Psychological Science
2016, Vol. 27(7) 1027–1035
© The Author(s) 2016
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0956797616645673
pss.sagepub.com
SAGE

The Power of the *Like* in Adolescence: Effects of Peer Influence on Neural and Behavioral Responses to Social Media

Lauren E. Sherman[1,2,3], Ashley A. Payton[4], Leanna M. Hernandez[2,4], Patricia M. Greenfield[1,3], and Mirella Dapretto[2,5]

[1]Department of Psychology, University of California, Los Angeles; [2]Ahmanson-Lovelace Brain Mapping Center, University of California, Los Angeles; [3]Children's Digital Media Center @ Los Angeles, University of California, Los Angeles, and California State University, Los Angeles; [4]Interdepartmental Neuroscience Program, University of California, Los Angeles; and [5]Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles

Abstract

We investigated a unique way in which adolescent peer influence occurs on social media. We developed a novel functional MRI (fMRI) paradigm to simulate Instagram, a popular social photo-sharing tool, and measured adolescents' behavioral and neural responses to *likes*, a quantifiable form of social endorsement and potential source of peer influence. Adolescents underwent fMRI while viewing photos ostensibly submitted to Instagram. They were more likely to *like* photos depicted with many likes than photos with few likes; this finding showed the influence of virtual peer endorsement and held for both neutral photos and photos of risky behaviors (e.g., drinking, smoking). Viewing photos with many (compared with few) likes was associated with greater activity in neural regions implicated in reward processing, social cognition, imitation, and attention. Furthermore, when adolescents viewed risky photos (as opposed to neutral photos), activation in the cognitive-control network decreased. These findings highlight possible mechanisms underlying peer influence during adolescence.

Keywords

adolescent development, social cognition, social influences, risk taking, neuroimaging, open materials

Received 9/12/15; Revision accepted 3/31/16

Social media are immensely popular among adolescents: Nearly 90% of American teens report being active users, and young people have continually outpaced other age groups in adopting new media (Lenhart, 2015). Given this prevalence, it is unsurprising that parents, educators, and the popular press have expressed concerns about the effects of social media on social-skill development and interpersonal interactions. Frequently, these concerns manifest themselves in questions about the effect of social media on the developing brain. Nonetheless, few studies have examined neural mechanisms underlying any kind of social-media use (Choudhury & McKinney, 2013; Mills, 2014).

The neural correlates of social-media use are particularly important to understand in the context of adolescence, and not only because adolescents are enthusiastic users. Adolescence is especially important for social cognitive development; it is theorized to be a sensitive period during which young people are uniquely attuned to the complexities of interpersonal relationships (Baird, 2012; Blakemore & Mills, 2014). Subcortical regions functionally associated with emotion processing and reward undergo considerable changes and reorganization during puberty (Brenhouse & Andersen, 2011; Sisk & Foster, 2004). The dopaminergic system and related regions in the striatum are implicated in

Corresponding Author:
Lauren E. Sherman, Department of Psychology, UCLA, 1285 Franz Hall, Box 951563, Los Angeles, CA 90095-1563
E-mail: lsherman@ucla.edu

potential mechanisms underlying two common features of adolescence: escalation in risk-taking behaviors and increased desire to spend time with and earn the approval of peers (Steinberg, 2008). For example, when adolescents completed a risky driving task alone or in the presence of peers, the presence of peers was associated with increases in both risk taking and activity in the nucleus accumbens (NAcc), a hub of reward circuitry (Chein, Albert, O'Brien, Uckert, & Steinberg, 2011). Smith, Chein, and Steinberg (2014) replicated these behavioral effects when peers were virtually connected, demonstrating that peer influence also occurs online (see also Cohen & Prinstein, 2006).

Less is known about how features unique to social media contribute to peer influence. For example, digital and in-person communication differ significantly in their affordance for quantifiable interactions. Whereas in-person communication is necessarily qualitative and involves subjective interpretation, many online environments allow for feedback that is purely quantitative. For example, a feature of most social media tools is the ability to *like* an image, text, or other piece of information, allowing for a simple, straightforward measure of peers' endorsement. For adolescents, who are particularly attuned to peer opinion, this *quantifiable social endorsement* may serve as a powerful motivator. Furthermore, quantifiable social endorsement provides a unique research opportunity: Although it is a form of interaction that occurs in the real world, it is simple enough to be experimentally manipulated.

The present study is, to our knowledge, the first to replicate social media interaction in the MRI scanner; however, important earlier work using behavioral and functional MRI (fMRI) methods has demonstrated how peer endorsement biases values (e.g., Campbell-Meiklejohn, Bach, Roepstorff, Dolan, & Frith, 2010; Izuma & Adolphs, 2013; Klucharev, Hytönen, Rijpkema, Smidts, & Fernández, 2009). In these studies, adults rated stimuli, then learned how other people rated the same stimuli, and finally rated the stimuli a second time. Participants changed their ratings to conform to those of peers or experts and showed greater NAcc activation during trials on which they agreed with these individuals than during trials on which they did not agree. Our study differs from previous work in that adolescents viewed content posted on social media simultaneously with information about its popularity—much as content is typically experienced online. We thus tested whether initial impressions were colored by the content's popularity and explored the overall effects of positive peer opinion on brain responses.

Specifically, we investigated the neural correlates of viewing photos with many or few *likes* to assess the role of quantifiable social endorsement in peer influence. We recruited adolescents to participate in an "internal social network" that simulated Instagram, a popular photo-sharing tool. Participants submitted their own Instagram photos, and they believed that all photos would be seen and liked by peers. We tested the possibility that the number of likes appearing under each photo would affect participants' responses. We hypothesized that participants would tend to like photos liked by more peers and refrain from liking less popular photos. We also hypothesized that neural responses to popular and unpopular photos would differ. Given previous research suggesting that peer presence heightens NAcc response (Chein et al., 2011), we predicted that viewing other people's photos that had a greater number of likes would similarly elicit greater NAcc activation. Evidence linking NAcc response to social evaluation (Meshi, Morawetz, & Heekeren, 2013) and sharing information about the self (Tamir & Mitchell, 2012), as well as the well-documented role of the NAcc in reward and reinforcement in general, suggests that viewing one's own popular photos would also elicit greater NAcc activity.

Peer influence is very important during adolescence; it is a means by which adolescents learn how to behave appropriately in their sociocultural environment. However, peer pressure can be maladaptive when it reinforces dangerous behaviors, such as drunk driving or drug use. Furthermore, young people frequently post content online depicting risky behaviors, and this may affect their peers' tendency to engage in such behaviors (Huang et al., 2014). Thus, we also investigated whether quantifiable social endorsement specifically influenced responses to risky behaviors by including photos depicting these behaviors. Well-established theories of adolescent risk taking suggest that the NAcc interacts with neural regions implicated in cognitive control during risky decision making (Casey, 2015; Steinberg, 2008). Accordingly, we directly compared adolescents' neural activity as they viewed risky images and neutral images to examine whether exposure to risky content online would influence activity in cognitive-control regions, regardless of the supposed popularity of the photos.

Method

Participants and fMRI paradigm

Thirty-four typically developing adolescents (18 female; age range = 13–18 years) participated in the present study. Two of these 34 participants were excluded from fMRI data analysis, 1 because of scan-console malfunction and 1 because of excessive motion. The sample size reflects the maximum number of participants that we were able to recruit given available funding, as well as timing constraints imposed by an institutional upgrade of the MRI magnet. Participants completed written consent in accordance with the institutional review board at the University of California, Los Angeles.

During recruitment, participants were informed that they would be involved in a study examining the brain's responses during social-media use. Participants were asked to submit photos from their own accounts on Instagram, a popular social-media tool used for sharing photos on mobile devices and the Internet. They were told that all of these photos would be combined to form an internal social network, that every participant would see a feed of these photos in the MRI scanner, and that the photos would appear as they did on Instagram. In reality, participants saw only some of their own photos while in the MRI scanner; all other stimuli were selected by the study team from among publicly available images on Instagram. During the laboratory visit, each participant was instructed that approximately 50 other adolescents had already viewed the feed of Instagram photos. This step was taken to establish the size of the audience, and to standardize how many likes would be regarded as many or few, irrespective of the size of a given participant's own social network. Participants were told that they could see how many times each photo was liked by previous participants and that the feed would be updated after their visit to reflect any new likes they contributed. In reality, the number of likes displayed under each image was assigned by the study team, as described later in this section.

The social-media task was presented to participants in the scanner using magnet-compatible 3-D goggles (VisuaStim; Resonance Technology, Inc., Northridge, CA) with a resolution of 800 × 640 pixels. The task mimicked the experience of browsing Instagram on a smartphone: Participants viewed a feed of photos, each of which was accompanied by text indicating how many other people had already liked the image. Photos were displayed one at a time on a white background accompanied by two on-screen buttons prompting the participant to choose "♥Like" to like the image or "→Next" to move on to the next image without liking it (Fig. 1). Images were presented for 3,000 ms, with an interstimulus interval that varied between 1,000 and 11,000 ms.

Participants saw 148 unique photos. These included 42 risky images and 66 neutral, nonrisky images. Risky photos depicted alcohol, cigarettes, marijuana, smoking paraphernalia, rude gestures, or adolescents (male and female) wearing provocative or skimpy clothing. Neutral photos depicted typical images (e.g., pictures of friends, food, and possessions) found on the social-media profiles of adolescents. Participants also saw 40 of the images they had submitted from their own Instagram accounts.

Across participants, all neutral and risky images were assigned both a popular value of 23 to 45 likes and an unpopular value of 0 to 22 likes. Two versions of the imaging paradigm were created: In Version 1, half of the photos in each category (risky, neutral) were displayed with a high number of likes and half were displayed with a low number of likes. In Version 2, the displayed popularity was opposite that in Version 1 (i.e., if a photo was displayed with many likes in Version 1, it was displayed with few likes in Version 2). Thus, half of the participants saw Version 1 of each image and half saw Version 2 of each image; this allowed us to hold the content and the aesthetic quality of the images constant while manipulating popularity.

To assign likes to participants' own images, author L. E. Sherman divided the 40 photos into groups on the basis of content (e.g., a people group or an objects group, depending on the participant). Then, each of the groups of photos was randomly split into two halves; one half was assigned many likes, and the other half was assigned few likes. Thus, the content of the popular and unpopular images was similar. Half of each participant's own photos appeared with 23 to 45 likes, and the other half appeared with 0 to 22 likes. Note that likes were not distributed continuously and evenly across the spectrum of 0 to 45. We did not expect neural and behavioral responses to vary linearly as the number of likes increased; instead, we hypothesized that participants would display qualitatively different responses to popular images than to unpopular images. Thus, we used a bimodal distribution of likes in which the majority was clustered between 30 and 45 likes (popular photos) or between 0 and 15 likes (unpopular photos). We chose to use a bimodal distribution to clearly differentiate popular and unpopular images. Of the 148 photos displayed during the scan, only 8 were depicted with intermediate values of 23 to 29 likes and 8 were depicted with 16 to 22 likes; these 16 images were included to avoid any suspicion among participants that might be caused by the obviously bimodal distribution. In light of our experimental manipulation, our categorical analyses reflect the difference between popular and unpopular images.

During the scan, participants were asked to view the images as they appeared and to decide whether they personally liked each image using the criteria they would normally use when deciding to like pictures on Instagram. Participants selected "♥Like" or "→Next" by pressing one of two buttons on a button box.

Data acquisition and analyses

Neuroimaging data were collected using a 3-T MRI scanner (Trio; Siemens Healthcare, Erlangen, Germany). The social-media paradigm was presented during a functional echoplanar, T2*-weighted gradient-echo scan lasting 11 min and 44 s (repetition time = 2,000 ms, echo time = 28 ms, flip angle = 90°, matrix size = 64 × 64, 34 axial slices, field of view = 192 mm, 4-mm slices with a 1-mm interslice gap). Button-press data were recorded in E-Prime

Fig. 1. Two examples of stimuli presented during the imaging paradigm. Participants saw innocuous photos of adolescents or everyday objects (e.g., the coffee drinks on the left) or images of objects related to risky behavior (e.g., the marijuana cigarette on the right) or adolescents engaging in risky behaviors. Images appeared as they would have in the Instagram app on a smartphone in the year 2014. The number of likes was displayed underneath each photo next to a heart icon, and the Instagram menu buttons were displayed beneath the likes. Finally, there were two buttons allowing participants to like an image ("♥Like") or to move on without liking the image ("➔Next").

(Version 2.0, Psychology Software Tools, Sharpsburg, PA) and converted to IBM SPSS Statistics format for analysis. Binomial tests were used to determine whether participants conformed to peers' responses more often than would be predicted by chance. fMRI data were preprocessed and analyzed using the Analysis of Functional NeuroImages (AFNI; Version 16.0.00) software environment (Cox, 1996) and the Functional MRI of the Brain software library (FSL; Jenkinson, Beckmann, Behrens, Woolrich, & Smith, 2012). Preprocessing for each individual's data included image realignment to correct for head motion, normalization to the standard stereotactic space of the Montreal Neurological Institute's (MNI) 152-brain template, and spatial smoothing using a 5-mm full-width, half-maximum Gaussian kernel to increase signal-to-noise ratio.

For each participant, linear contrasts were calculated for several planned comparisons. Specifically, we modeled three linear contrasts comparing popular photos (many likes) and unpopular photos (few likes) in all three categories (i.e., neutral photos, risky photos, and participants' photos). In addition to modeling the six types of stimuli at the first level, we included several other parameters. These included the participant's button-press choice and reaction time for each trial and the luminosity of each image as determined using Adobe Photoshop. Group-level random-effects analyses were then conducted across all participants. At the group level, a prethreshold binary mask consisting of all regions exhibiting significant activity for any type of photo, compared with a fixation cross on a white background, was used to restrict our analyses to regions displaying significant task-related activity. Specifically, we first individually contrasted the six types of stimuli (e.g., neutral photos with many likes, neutral photos with few likes, risky photos with many likes) > fixation and then added the maps of each of these individual contrasts (thresholded at $z > 1.7$, corrected for multiple comparisons at $p < .05$) together. The final mask covered a considerable portion of the cortex and subcortex. Along with all of our group contrast maps, it is available for download at NeuroVault (http://neurovault.org/collections/RYSBTTMN/). We performed contrasts examining the effect of popularity (many likes > few likes and the reverse) for neutral photos, risky photos, and participants' photos. We also compared all neutral photos ostensibly submitted by peers with all risky photos ostensibly submitted by peers.

To test our a priori hypothesis that popular photos would elicit significantly greater activation in the bilateral

NAcc than unpopular photos would, we used a small-volume-correction approach. Our functional regions of interest (ROIs), derived from an independent sample of participants completing a monetary-incentive-delay task (Tamir & Mitchell, 2012), consisted of two 8-mm spheres in the left and right NAcc (MNI coordinates: $x = 10$, $y = 6$, $z = -4$, and $x = -8$, $y = 4$, $z = -6$, respectively). AFNI's 3dClustSim was used to determine that a contiguous cluster of 53 or more voxels was necessary to meet statistical criteria within these ROIs. To examine whether the many likes > few likes contrast differed significantly as a function of type of photo (neutral, risky, participant), we extracted parameter estimates (regression coefficients) from the bilateral ROIs for each contrast of interest and performed paired-samples t tests using IBM SPSS.

Results

To determine whether participants were significantly more likely than chance to match the supposed opinions of peers (i.e., to like popular images and to refrain from liking unpopular images), we conducted a series of binomial tests. Across all photos presented during the scan, participants matched their peers significantly more frequently than expected by chance ($p < .00001$). This effect was also significant for each individual type of photo, including neutral images ostensibly provided by peers ($p = .03$), images depicting risk-taking behaviors ostensibly provided by peers ($p = .03$), and the participants' own images ($p < .00001$). The effect was significantly larger for participants' own photos than for either neutral images, $\chi^2(1, N = 3{,}544) = 10.1$, $p = .001$, or risky images, $\chi^2(1, N = 2{,}736) = 6.6$, $p = .01$.

Neural responses also differed according to the number of likes for neutral, risky, and participants' own photos. Figure 2a depicts regions in which activity was significantly greater when photos were depicted as having garnered many versus few likes for neutral, risky, and participants' own photos. The regions of significantly greater activity for many likes compared with few likes differed by photo type. When participants viewed neutral photos with many likes, they showed significantly greater activity in the visual cortex extending to the precuneus and in the cerebellum (see Table S1 in the Supplemental Material available online). When participants viewed risky photos with many likes (compared with risky photos with few likes), significantly greater activity was found in one cluster in the left frontal cortex, extending from the precentral gyrus through the middle frontal gyrus and inferior frontal gyrus (Table S1). When participants viewed their own photos, significantly greater activity in response to photos with many likes (compared with photos with few likes) was observed in several regions (Table S1). These included areas implicated in social cognition, such as the precuneus, medial prefrontal cortex, left temporal pole, lateral occipital cortex, and hippocampus (Mars et al., 2012; Zaki & Ochsner, 2009), as well as reward learning and motivation, including the nucleus accumbens, caudate, putamen, thalamus, ventral tegmental area, and brain stem (e.g., Haruno & Kawato, 2006; Schott et al., 2008).[1] Table S1 includes a complete list of regions. For all three photo types, the reverse contrast (few likes > many likes) yielded no significant activation in the whole brain.

Neural responses also differed according to whether the photo depicted risky behavior (Fig. 2b). When participants viewed neutral images (compared with risky images) ostensibly submitted by peers, significantly greater activity was observed in bilateral occipital cortex, medial prefrontal cortex, and the inferior frontal gyrus (for a complete list of regions, see Table S2 in the Supplemental Material). When viewing risky images compared with neutral images (i.e., the reverse contrast), participants demonstrated significantly less activation in a network of regions implicated in cognitive control and response inhibition (e.g., Blasi et al., 2006; Bressler & Menon, 2010; Sherman et al., 2014), including dorsal anterior cingulate cortex, bilateral prefrontal cortex, and lateral parietal cortex (Table S2).[2]

In addition to whole-brain analyses, we conducted ROI analyses on the basis of our a priori hypothesis that photos depicted with many likes would elicit significantly greater activation in the bilateral NAcc than would those depicted with few likes. Consistent with our hypothesis, there was greater activity in the left NAcc when participants viewed neutral images that had many likes than when they viewed neutral images that had few likes. We also observed greater bilateral NAcc activation when participants viewed their own images for the many likes > few likes contrast. For images depicting risk-taking behavior, likes had no effect on brain response in the NAcc ROI. In the right NAcc, activation was significantly greater when participants viewed their own photos than when viewing other people's neutral images, $t(31) = 2.34$, $p = .026$, or risky images, $t(31) = 2.45$, $p = .02$, but did not differ significantly in the left NAcc (for all comparisons, $p > .10$).

Discussion

The present study highlights a new and unique way in which peer influence occurs on social media: through quantifiable social endorsement. We found that the popularity of a photo had a significant effect on the way that photo was perceived. Adolescents were more likely to like a photo—even one portraying risky behaviors, such as smoking marijuana or drinking alcohol—if that photo had received more likes from peers. This effect was

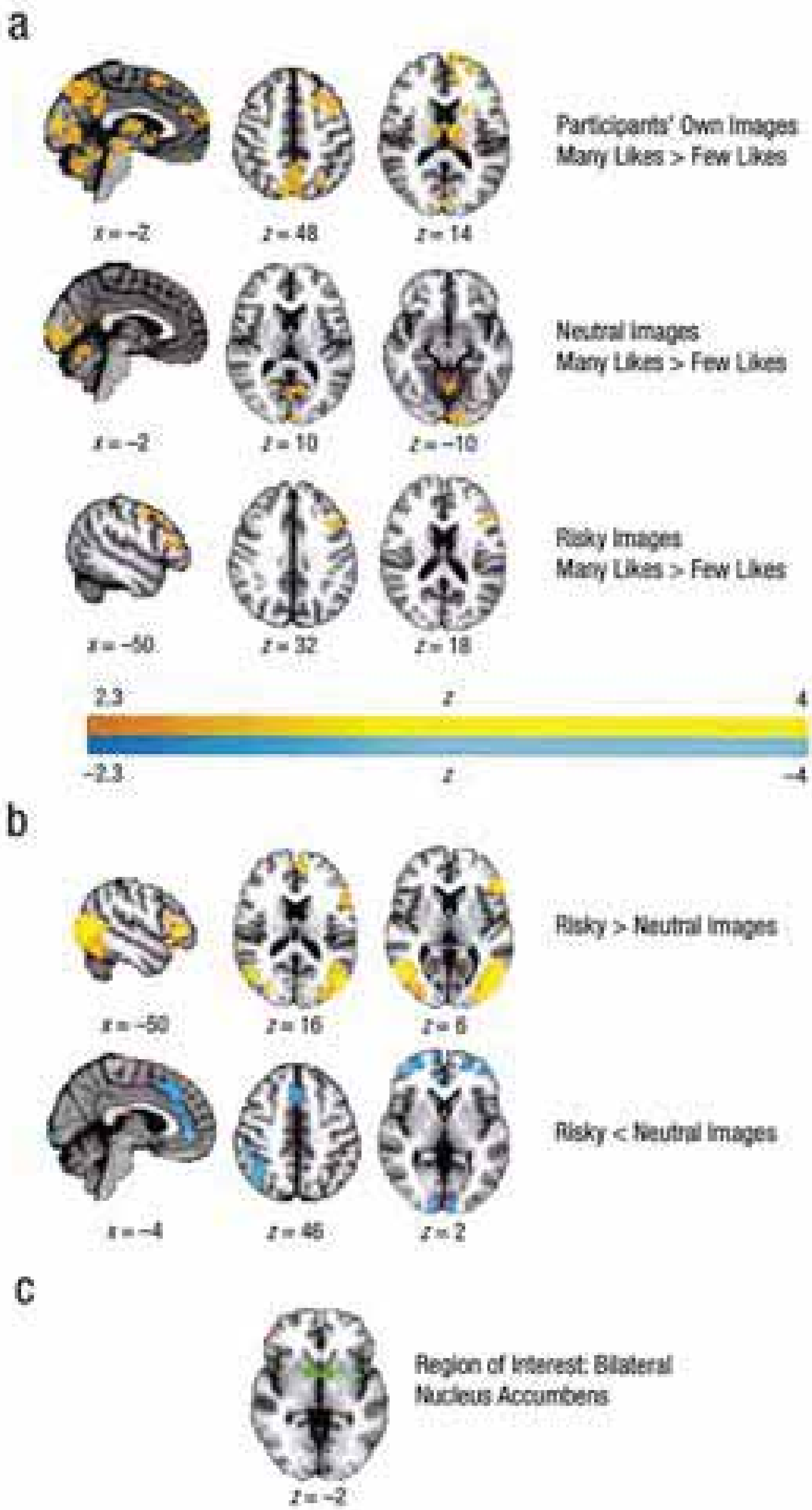


Fig. 2. Neural responses to Instagram photos with many likes compared with photos with few likes. The brain maps in (a) show neural regions with significant activity ($z > 2.3$, cluster corrected at $p < .05$) for the many likes > few likes contrast, for each of the three types of photos. The brain maps in (b) show neural regions with significant activity ($z > 2.3$, cluster corrected at $p < .05$) for the risky > neutral contrast and the risky < neutral contrast. Brain images are shown by radiological convention (i.e., left side of the brain is on the right). The brain map in (c) shows the location of the region of interest in the nucleus accumbens that was identified using a monetary-incentive-delay task in an independent sample of young adults (Tamir & Mitchell, 2012).

especially strong for photos the participants themselves had supplied. Adolescence is a period during which self-presentation is particularly important, including on social media; thus, this significantly greater effect may reflect the relative importance of self-presentation versus providing feedback to others.

Neural responses also differed according to number of likes. For all three types of photos, participants exhibited greater brain activity for photos with more likes. The regions of greater activity included areas implicated in social cognition and social memories, including the precuneus, medial prefrontal cortex, and hippocampus (Mars et al., 2012; Zaki & Ochsner, 2009), as well as the inferior frontal gyrus, which is implicated in imitation (Pfeifer, Iacoboni, Mazziotta, & Dapretto, 2008). When participants viewed their own photographs or neutral photographs ostensibly submitted by peers, greater activity in the visual cortex was observed in response to photos with many likes compared with photos with few likes, even though we controlled for photos' luminosity and content. The increased activation suggests that participants may have scanned popular images with greater care. Taken together, our imaging findings suggest that adolescents perceive information online in a qualitatively different way when they believe that this information is valued more highly by peers. The exact nature of these changes differs depending on the content depicted in the photo.

Our ROI analysis suggests that the NAcc, an important hub of the brain's reward circuitry, is implicated in the experience of receiving positive feedback on one's own images as well as viewing other people's images that have been endorsed by peers. The NAcc response, like our behavioral effects, was particularly robust for participants' own photos, suggesting that self-presentation can be especially rewarding and a motivation for using social networks (Manago, Graham, Greenfield, & Salimkhan, 2008). The popularity of risky photos (or lack thereof) had no differential effect on NAcc response. However, several participants in our adolescent sample reported no experiences with drugs and alcohol; this lack of familiarity may have contributed to the failure to detect a peer effect in the NAcc when comparing popular and unpopular risky images. Future research should examine the effect of popularity on NAcc response to risky photos in adolescents who report greater experience with drugs and alcohol.

Although quantifiable social endorsement is a relatively new phenomenon, we believe that the implications of this experiment extend beyond the digital context. Quantifiable social endorsement is a simple but nonetheless significant example of sociocultural learning; a like is a social cue specific to adolescents' cultural sphere, and adolescents use this cue to learn how to navigate their social world. Adolescents learn from quantifiable social endorsement in multiple ways, as evidenced by participants' differentiated neural responses to their own and other people's photos. Peers socialize one another to norms in multiple modes, including modeling appropriate behavior (behavioral display) and reinforcing appropriate behavior in other people (behavioral reinforcement; Brown, Bakken, Ameringer, & Mahon, 2008). Social media embody both modes of socialization: Adolescents model appropriate behavior and interests through the images they post (behavioral display) and reinforce peers' behavior through the provision of likes (behavioral reinforcement). Unlike offline forms of peer influence, however, quantifiable social endorsement is straightforward, unambiguous, and, as the name suggests, purely quantitative.

Although the present study does not allow us to directly compare in-person versus online peer influence, our findings are in line with results from previous research suggesting that the presence of peers heightens responses in reward circuitry and leads to differences in behavioral decision making (Chein et al., 2011). Furthermore, the present inquiry is, to our knowledge, the first to document that quantifiable social endorsement, a ubiquitous feature of social media, produces these measurable neural and behavioral effects. Future research should build on our findings to investigate how individual differences in neural response map onto behavioral outcomes: Can individual neural responses predict the degree of conformity that adolescents will demonstrate?

Sociocultural learning can be adaptive, in that it allows adolescents to flexibly learn from their environment. In the case of socialization to risky behavior, however, it can also be maladaptive. Multiple theoretical models (Casey, 2015; Steinberg, 2008) posit that risk taking in adolescence arises in part from heightened neural sensitivity to reward combined with immature capacity for cognitive control. In results that are in line with these models, we found that a network implicated in cognitive control (e.g., Seeley et al., 2007) was less active when participants viewed images depicting risky behavior (compared with neutral images). Certainly, viewing photos online does not, in itself, constitute a risk. It is therefore all the more striking that when simply viewing photos of risky behaviors ostensibly taken and posted by peers, adolescents exhibited decreased activation of the cognitive control network, possibly reflecting a mechanism by which peer behaviors disinhibit cognitive control in high-risk scenarios, thereby increasing the likelihood of engaging in risk taking. Future research should examine whether this decreased activation occurs into adulthood as well, or if this finding potentially reflects the immaturity of the prefrontal cortex in adolescence. Likewise, future research can shed light on whether the NAcc response to

social reward shown in the present study is particularly heightened in adolescence, in line with previous research on monetary reward (Braams, van Duijvenvoorde, Peper, & Crone, 2015).

Our findings and approach have implications not only for social media researchers, but also for those studying social cognition more broadly. Social media provide a compelling opportunity to examine social interaction in an ecologically valid context. Typically, in the confines of an MRI scanner, social interaction is limited and artificial. Because social media exist on a screen, however, they can be effectively imported into the scanner environment. Our study provides proof of concept for quantifiable social endorsement, a ubiquitous form of online interaction that is easily experimentally manipulated. Future research can build on this foundation to examine how neural responses to quantifiable social endorsement predict individual differences in a variety of behavioral and psychological domains.

Action Editor

Eddie Harmon-Jones served as action editor for this article.

Author Contributions

L. E. Sherman developed the study concept, and L. E. Sherman, M. Dapretto, and P. M. Greenfield contributed to the study design. Data collection was performed by L. E. Sherman, A. A. Payton, and L. M. Hernandez. L. E. Sherman and A. A. Payton performed the data analysis and interpretation under the supervision of M. Dapretto and P. M. Greenfield. L. E. Sherman drafted the manuscript, and M. Dapretto and P. M. Greenfield provided important revisions. All the authors approved the final version of the manuscript for submission.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

This research was supported by Grants C06-RR012169 and C06-RR015431 from the National Center for Research Resources, by Grant S10-OD011939 from the Office of the Director of the National Institutes of Health (NIH), by National Institute on Drug Abuse National Research Service Award F31-DA038578-01A1 (to L. E. Sherman), and by Brain Mapping Medical Research Organization, Brain Mapping Support Foundation, Pierson-Lovelace Foundation, The Ahmanson Foundation, Capital Group Companies Charitable Foundation, William M. and Linda R. Dietel Philanthropic Fund, and Northstar Fund. Authors are solely responsible for the content, which may not represent the official views of NIH.

Supplemental Material

Additional supporting information can be found at http://pss.sagepub.com/content/by/supplemental-data

Open Practices

All materials have been made publicly available via Open Science Framework and can be accessed at https://osf.io/atjxd. The complete Open Practices Disclosure for this article can be found at http://pss.sagepub.com/content/by/supplemental-data. This article has received the badge for Open Materials. More information about the Open Practices badges can be found at https://osf.io/tvyxz/wiki/1.%20View%20the%20Badges/ and http://pss.sagepub.com/content/25/1/3.full.

Notes

1. The first set of regions also resembled the map for the term "social" on Neurosynth (http://neurosynth.org; a large-scale database of neuroimaging studies that provides meta-analytic reverse-inference analyses) as of January 2016 (Yarkoni et al., 2011). The second set of regions also resembled the map for the term "reward" on Neurosynth as of January 2016.

2. This set of regions also resembled the Neurosynth map for the term "cognitive control" as of January 2016.

References

Baird, A. A. (2012). The terrible twelves. In P. D. Zelazo, M. Chandler, & E. Crone (Eds.), *Developmental social cognitive neuroscience* (pp. 191–207). New York, NY: Psychology Press.

Blakemore, S. J., & Mills, K. L. (2014). Is adolescence a sensitive period for sociocultural processing? *Annual Review of Psychology, 65*, 187–207.

Blasi, G., Goldberg, T. E., Weickert, T., Das, S., Kohn, P., Zoltick, B., . . . Mattay, V. S. (2006). Brain regions underlying response inhibition and interference monitoring and suppression. *European Journal of Neuroscience, 23*, 1658–1664.

Braams, B. R., van Duijvenvoorde, A. C., Peper, J. S., & Crone, E. A. (2015). Longitudinal changes in adolescent risk-taking: A comprehensive study of neural responses to rewards, pubertal development, and risk-taking behavior. *The Journal of Neuroscience, 35*, 7226–7238.

Brenhouse, H. C., & Andersen, S. L. (2011). Developmental trajectories during adolescence in males and females: A cross-species understanding of underlying brain changes. *Neuroscience & Biobehavioral Reviews, 35*, 1687–1703.

Bressler, S. L., & Menon, V. (2010). Large-scale brain networks in cognition: Emerging methods and principles. *Trends in Cognitive Sciences, 14*, 277–290.

Brown, B. B., Bakken, J. P., Ameringer, S. W., & Mahon, S. D. (2008). *A comprehensive conceptualization of the peer influence process in adolescence*. New York, NY: Guilford Press.

Campbell-Meiklejohn, D. K., Bach, D. R., Roepstorff, A., Dolan, R. J., & Frith, C. D. (2010). How the opinion of others affects our valuation of objects. *Current Biology, 20*, 1165–1170.

Casey, B. J. (2015). Beyond simple models of self-control to circuit-based accounts of adolescent behavior. *Annual Review of Psychology, 66*, 295–319.

Chein, J., Albert, D., O'Brien, L., Uckert, K., & Steinberg, L. (2011). Peers increase adolescent risk taking by enhancing activity in the brain's reward circuitry. *Developmental Science, 14*(2), F1–F10.

Choudhury, S., & McKinney, K. A. (2013). Digital media, the developing brain and the interpretive plasticity of neuroplasticity. *Transcultural Psychiatry, 50*, 192–215.

Cohen, G. L., & Prinstein, M. J. (2006). Peer contagion of aggression and health risk behavior among adolescent males: An experimental investigation of effects on public conduct and private attitudes. *Child Development, 77*, 967–983.

Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research, 29*, 162–173.

Haruno, M., & Kawato, M. (2006). Heterarchical reinforcement-learning model for integration of multiple cortico-striatal loops: fMRI examination in stimulus-action-reward association learning. *Neural Networks, 19*, 1242–1254.

Huang, G. C., Unger, J. B., Soto, D., Fujimoto, K., Pentz, M. A., Jordan-Marsh, M., & Valente, T. W. (2014). Peer influences: The impact of online and offline friendship networks on adolescent smoking and alcohol use. *Journal of Adolescent Health, 54*, 508–514.

Izuma, K., & Adolphs, R. (2013). Social manipulation of preference in the human brain. *Neuron, 78*, 563–573.

Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., & Smith, S. M. (2012). FSL. *NeuroImage, 62*, 782–790.

Klucharev, V., Hytönen, K., Rijpkema, M., Smidts, A., & Fernández, G. (2009). Reinforcement learning signal predicts social conformity. *Neuron, 61*, 140–151.

Lenhart, A. (2015). *Teens, social media & technology overview 2015*. Retrieved from the Pew Research Center Web site: http://www.pewinternet.org/2015/04/09/teens-social-media-technology-2015/

Manago, A. M., Graham, M. B., Greenfield, P. M., & Salimkhan, G. (2008). Self-presentation and gender on MySpace. *Journal of Applied Developmental Psychology, 29*, 446–458.

Mars, R. B., Neubert, F., Noonan, M. P., Sallet, J., Toni, I., & Rushworth, M. F. S. (2012). On the relationship between the "default mode network" and the "social brain." *Frontiers in Human Neuroscience, 6*, Article 189. doi: 10.3389/fnhum.2012.00189

Meshi, D., Morawetz, C., & Heekeren, H. R. (2013). Nucleus accumbens response to gains in reputation for the self relative to gains for others predicts social media use. *Frontiers in Human Neuroscience, 7*, Article 439. doi:10.3389/fnhum.2013.00439

Mills, K. L. (2014). Effects of Internet use on the adolescent brain: Despite popular claims, experimental evidence remains scarce. *Trends in Cognitive Sciences, 18*, 385–387.

Pfeifer, J. H., Iacoboni, M., Mazziotta, J. C., & Dapretto, M. (2008). Mirroring others' emotions relates to empathy and interpersonal competence in children. *NeuroImage, 39*, 2076–2085.

Schott, B. H., Minuzzi, L., Krebs, R. M., Elmenhorst, D., Lang, M., Winz, O. H., . . . Düzel, E. (2008). Mesolimbic functional magnetic resonance imaging activations during reward anticipation correlate with reward-related ventral striatal dopamine release. *The Journal of Neuroscience, 28*, 14311–14319.

Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., . . . Greicius, M. D. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. *The Journal of Neuroscience, 27*, 2349–2356.

Sherman, L. E., Rudie, J. D., Pfeifer, J. H., Masten, C. L., McNealy, K., & Dapretto, M. (2014). Development of the Default Mode and Central Executive Networks across early adolescence: A longitudinal study. *Developmental Cognitive Neuroscience, 10*, 148–159.

Sisk, C. L., & Foster, D. L. (2004). The neural basis of puberty and adolescence. *Nature Neuroscience, 7*, 1040–1042.

Smith, A. R., Chein, J., & Steinberg, L. (2014). Peers increase adolescent risk taking even when the probabilities of negative outcomes are known. *Developmental Psychology, 50*, 1564–1568.

Steinberg, L. (2008). A social neuroscience perspective on adolescent risk-taking. *Developmental Review, 28*, 78–106.

Tamir, D. I., & Mitchell, J. P. (2012). Disclosing information about the self is intrinsically rewarding. *Proceedings of the National Academy of Sciences, USA, 109*, 8038–8043.

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods, 8*, 665–670.

Zaki, J., & Ochsner, K. (2009). The need for a cognitive neuroscience of naturalistic social cognition. *Annals of the New York Academy of Sciences, 1667*, 16–30.

4

Telematics and Informatics Reports 11 (2023) 100087

Contents lists available at ScienceDirect

Telematics and Informatics Reports

journal homepage: www.elsevier.com/locate/teler

Snapchat streaks—How are these forms of gamified interactions associated with problematic smartphone use and fear of missing out among early adolescents?

Christina M. van Essen [a,*], Joris Van Ouytsel [a,b]

[a] Hugh Downs School of Human Communication, Arizona State University, Tempe, AZ 85287, United States
[b] Department of Communication Studies, University of Antwerp, 2000 Antwerp, Belgium

Streaks

ARTICLE INFO

Keywords:
Snapchat streaks
Adolescents
Social media

ABSTRACT

Snapchat offers a unique function, the Snapchat Streak, which is a gamified function within the app that motivates users to participate in daily interactions. This feature of the application can aid users in building a friendship with their peers. Given the requirement of interacting on the platform every 24 hours, our exploratory study aims to investigate how Snapchat streaks are associated with Fear of Missing Out (FOMO), problematic smartphone use and social media self-control. W
study among a final sample of 2483 early adolescents (M_{age} = 12.46 years old
ing community of Belgium. The results indicate that the girls w
streak and were more likely to engage in streaks
ated with the engagement in Snapchat streaks.
f-control were correlated with the number of
eaks with, albeit it being a weak relationship.
use are provided.

Streaks

Introduction

Over the past years, Snapchat has become
social media applications among adolescents.
from Belgium, the country where this research
the youth between the ages of 12 and 18 yea
account [1]. Snapchat allows its users to send
pictures and videos to others. The content sent ca
or twice for a limited amount of time (between 1
the timer expires, the images disappear permane
still have the option to take a screenshot of the co
has a "My Story" feature where users can post im
publish the content for 24 hours after which it disap
are visible to all the Snapchat contacts [2].

Prior work has found that many adolescents perce
appealing medium to communicate with their frien
especially because the disappearing images can remain outside of parental control. Although research on adolescents' motivations for using Snapchat is limited, a study among a convenience sample of young adults has found that the most frequently sent and received content on Snapchat consists

50% of the participants, followed by
ost common recipient of the message
nting for 55% and 18% respectively
's simple design and use it to share
hat they would not share on other

of the content, Snapchat users tend
ul content. They perceive Snapchat
re casual photo sharing compared
Facebook or Instagram. [6,7,8].
, one of the main motivations to
porn, along with various other
contrast with the use of Instagram, where
curated versions of their lives, by carefully selecting the images and by being more intentional about the long-term nature of the communication that they have through the app [7,9]. For instance, a study investigating users who post food images, commonly known as "food porn," found that these individuals had diverse reasons for sharing such photos. Instagram was primarily employed for public sharing (e.g., for the purpose of engaging in self-presentation), while Snapchat was

* Corresponding author at: Hugh Downs School of Human Communication, Arizona State University, 950 S Forest Mall, Tempe, AZ 85287-1205, United States.
E-mail addresses: [illegible] (C.M. van Essen), [illegible] (J. Van Ouytsel).

https://doi.org/10.1016/j.teler.2023.100087
Received 12 September 2022; Received in revised form 5 July 2023; Accepted 27 July 2023
Available online 28 July 2023

Telematics and Informatics Reports 11 (2023) 100087

Contents lists available at ScienceDirect

Telematics and Informatics Reports

journal homepage: www.elsevier.com/locate/teler

Snapchat streaks—How are these forms of gamified interactions associated with problematic smartphone use and fear of missing out among early adolescents?

Christina M. van Essen [a,*], Joris Van Ouytsel [a,b]

[a] Hugh Downs School of Human Communication, Arizona State University, Tempe, AZ 85287, United States
[b] Department of Communication Studies, University of Antwerp, 2000 Antwerp, Belgium

ARTICLE INFO

Keywords:
Snapchat streaks
Adolescents
Social media

ABSTRACT

Snapchat offers a unique function, the Snapchat Streak, which is a gamified function within the app that motivates users to participate in daily interactions. This feature of the application can aid users in building a friendship with their peers. Given the requirement of interacting on the platform every 24 hours, our exploratory study aims to investigate how Snapchat streaks are associated with Fear of Missing Out (FOMO), problematic smartphone use and social media self-control. We conducted a study among a final sample of 2483 early adolescents (M_{age} = 13.46 years old; SD = 0.894) in the Dutch-speaking community of Belgium. The results indicate that the girls were more likely than boys to engage in a Snapchat streak and were more likely to engage in streaks for a longer period of time. Problematic smartphone use was associated with the engagement in Snapchat streaks. Lastly, FOMO, problematic smartphone use, and social media self-control were correlated with the number of people and the number of days adolescents maintained Snapchat streaks with, albeit it being a weak relationship. Implications of the findings for understanding adolescent Snapchat use are provided.

Introduction

Over the past years, Snapchat has become one of the most popular social media applications among adolescents. According to recent data from Belgium, the country where this research was carried out, 91% of the youth between the ages of 12 and 18 years old owns a Snapchat account [1]. Snapchat allows its users to send private text messages, pictures and videos to others. The content sent can only be opened once or twice for a limited amount of time (between 1 and 10 seconds). Once the timer expires, the images disappear permanently. Of course, users still have the option to take a screenshot of the content. Snapchat also has a "My Story" feature where users can post images and videos and publish the content for 24 hours after which it disappears. These stories are visible to all the Snapchat contacts [2].

Prior work has found that many adolescents perceive Snapchat as an appealing medium to communicate with their friends [3], especially because the disappearing images can remain outside of parental control. Although research on adolescents' motivations for using Snapchat is limited, a study among a convenience sample of young adults has found that the most frequently sent and received content on Snapchat consists of selfies, which was reported by 50% of the participants, followed by screenshots at only 7% [4]. The most common recipient of the message is a close friend or partner, accounting for 55% and 18% respectively [4]. Young adults value Snapchat's simple design and use it to share personal or entertaining content that they would not share on other platforms [4,5].

Because of the temporary nature of the content, Snapchat users tend to share mundane, humorous or playful content. They perceive Snapchat as a platform for less curated and more casual photo sharing compared to other social media platforms like Facebook or Instagram. [6,7,8]. Unsurprisingly, among college students, one of the main motivations to use Snapchat is for entertainment purposes, along with various other motivations [9]. This is in contrast with the use of Instagram, where users share curated versions of their lives, by carefully selecting the images and by being more intentional about the long-term nature of the communication that they have through the app [7,5]. For instance, a study investigating users who post food images, commonly known as "food porn," found that these individuals had diverse reasons for sharing such photos. Instagram was primarily employed for public sharing (e.g., for the purpose of engaging in self-presentation), while Snapchat was

* Corresponding author at: Hugh Downs School of Human Communication, Arizona State University, 950 S Forest Mall, Tempe, AZ 85287-1205, United States.
E-mail addresses: [illegible] (C.M. van Essen), [illegible] (J. Van Ouytsel).

https://doi.org/10.1016/j.teler.2023.100087
Received 12 September 2022; Received in revised form 5 July 2023; Accepted 27 July 2023
Available online 28 July 2023

used to share private images (e.g., to connect with friends and family members). This emphasizes the distinct motivations and affordances of the ephemeral nature of Snapchat communication [10].

The ephemeral nature of the content on Snapchat and its nature that allows for the sending of spontaneous and less curated content, also raised substantial concern for the application's ability to be used for the purposes of harassment, sexting and sexual solicitation (e.g., [11,12]). Qualitative studies have found that adolescents engage in sexting through Snapchat, taking advantage of the platform's disappearing images feature (e.g., [13,14]). This behavior introduces the risk of images being disseminated to wider audiences without permission of the creator and can have considerable long-term consequences [15].

Adolescent Snapchat users also experience image-based sexual abuse (i.e., through receiving unsolicited sexual images), inappropriate messages and other forms of harassment [7,11]. Adult Snapchat users employ the app for seeking sexual access, requesting hookups, or soliciting sexting images. Notably, men were found to be more likely than women to use Snapchat for requesting hookups or sexts [16]. The disappearing nature of the images also makes the application an attractive medium for perpetrators of online grooming, as Snapchat has been used in many cases of child sexual abuse to gain access to the victim and to coerce the victim through online sexual abuse [17].

Snapchat streaks

Snapchat has several features that encourages its users to frequently communicate with each other. Snapchat has several built-in gamification features that encourage frequent use, such as badges, points, or challenges [18,19]. One example is the Snapchat score which ranks users that use the application most often [20]. Another popular feature is the so-called "Snapchat Streak" [21]. During a Snapchat streak, users send each other at least one picture every 24 hours. If they do so for more than three days in a row, a flame icon will appear next to the name of the friend. The number next to the flame indicates how many days in a row the users have exchanged the images with each other. If users fail to send a back-and-forth message within 24 hours, they will lose the streak and may have to start over again [18,20].

Maintaining a Snapchat streak can become an important goal for adolescents [21]. In focus group interviews conducted by Hristova & Lieberoth [21], adolescents indicated that engaging in a Snapchat streak offers them the opportunity to renegotiate their relationship with their friends and to participate in a shared project. Additionally, the study revealed that students viewed breaking a Snapchat streak as a personal rejection [21]. Snapchat streaks, thus, serve as a way for teenagers to evaluate the strength of their interpersonal connections. Given the perceived importance of Snapchat streaks within their friendships, adolescents often engage in extensive communication to maintain their streaks, such as reminding their partners to snap before the 24-hour deadline in order to rescue a streak, or by asking for an explanation when a Snapchat streak was lost [22,21]. Studies in Austria and Belgium found that adolescents will share the passwords of their Snapchat accounts so that others can maintain their streaks if they are unable to do so, during times in which they do not have access to their smartphones [19,23]. Qualitative studies found that the relational expectations associated with Snapchat streaks, as well as the frequent notifications they generate, can lead to feelings of stress among early adolescents. [22,24].

So-called friend emojis add to the perceptions of social pressure when engaging in Snapchat streaks. By frequently direct messaging individuals, one can establish an online 'Best Friends Forever' (BFF) relationship [25] where one can be a "Super BFF" with someone when both parties have sent each other the most Snapchats daily for two months in a row. Unlike other social media platforms, the frequency of communication on the platform, and thus the strength of the relationships that users have on the platform is open for others to see, while at the same time the content of the Snapchat messages disappears. This function might lead to distress or jealousy in adolescent friendships and even their romantic relationships [26,27].

These gamification features of social media applications, such as Snapchat streaks, have become the cause of growing public concern. Unsurprisingly, Snapchat streaks have become part of the larger risk discourse and moral panics that surround teenagers' media use [28]. In the popular press, they are perceived to be related to problematic smartphone use, smartphone addiction, cyberbullying victimization and online 'stranger danger'. For example, the British newspaper 'The Sun' published an article entitled: "Snapchat Streaks 'should be stopped' claim worried parents who fear it puts teens at risk", in which these fears were laid out [29]. In the United States, the Republican Senator Josh Hawley introduced a bill that, amongst others, aimed to restrict social media applications to give awards to users for higher engagement [30]. The press release that accompanied the bill and the subsequent news coverage of the legislation specifically mentioned Snapchat streaks as one of the features that were targeted by this legislation [31,32]. Given the moral panics and risk discourse related to Snapchat streaks, our study aims to examine the associations between engagement in Snapchat streaks, and some measures of media use, including Fear of Missing Out (FOMO), problematic smartphone use and social media self-control.

Fear of Missing Out (FOMO)

Prior work has focused more broadly on examining the role of Fear Of Missing Out (FOMO) in adolescents' social media use. FOMO can be broadly defined as "a desire to be online and a constant urge to check social media" ([33], p.1). A Belgian study found that almost one in ten of the adolescents in their study agreed with the items that assessed FOMO [34]. Franchina et al. [35] conducted a separate study which discovered that Belgian teenagers, on average, scored neutral on their levels of FOMO, but girls reported higher levels compared to boys. The study did not find a significant relationship between age or school track and experiences of FOMO. Throuvala et al. [5] found that FOMO is a key motivator for teenagers to utilize social media, as they aspire to be included in and informed about their peers' endeavors. When adolescents can continuously monitor the whereabouts and social engagement of their peers, they may be more likely to engage in a constant social comparison between themselves and their peers that can lead to significant stressors. For example, another Belgian study [34] suggests that adolescents experienced increased stress about popularity and belonging when their FOMO increased.

The possibility of increased social media usage due to FOMO can be concerning to parents, in particular because they already express worry over their children's excessive social media consumption [36]. Parental concerns about such excessive use are not unwarranted considering research conducted in Turkey showed that 27.5% of variance in problematic phone use was related to the youth's FOMO [37]. Our study aims to investigate the potential relationship between FOMO and Snapchat streaks, as the daily gamified social interactions involved may lead to feelings of social pressure among users to maintain the streaks. Moreover, users may perceive a sense of missing out on social connections with their friends and being aware of their social experiences, including conversations and shared images, if they fail to maintain a Snapchat streak with their friends. Users who do not maintain a Snapchat streak may feel like they are missing out shared experiences.

Problematic smartphone use

Where FOMO focuses on the unease caused by the idea to miss out on seemingly important information, connections, or interactions, problematic smartphone use addresses a deeper-rooted issue in which smartphone use will take precedence in importance over face-to-face interactions and relationships. Although research is inconclusive on whether problematic smartphone use may be considered an addiction (e.g. [38,39]), the obsessive use of smartphones can become an issue in

the formation and maintenance of the adolescent-parent relationships [40,41]. Parents frequently fear the impact of social technologies on the development of the social interactions of their children, in fact, [42] address that the fear of lack of presence in "real life" and the fear that mobile phones create a digital divide between parents and children as two of the seven fears in relation to adolescents' media use.

Social media self control

A major component to the frequency of use of social media is the ability to resist using it. Du et al. [43] established the social media self-control failure scale (SMSCF), and defined social media self-control failure as how frequently "social media users give in to the desire to use social media, even though its use at that moment conflicts with other goals, makes them use their time less efficiently, and delays other things they want or need to do" ([43]; p. 74). This measure serves to bridge the gap between the inability to classify internet addiction in accordance with the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) [43] and the need to label the potential obsessive nature of internet use.

The gamification features added to many may contribute to the extensive use of social media. The impact of gamification has shown potential impact on adolescent behavior (e.g. [44,45]). With the strict 24-hours time restrictions Snapchat uses to maintain their streaks, some adolescents may start prioritizing the use of the platform over any other responsibilities and social interactions they can partake in. This in particular will drive adolescents who already struggle in their inability to control their use of social media to partake in such gamified maintenance of relationships.

The present study

Despite the popularity of the Snapchat streak feature among early adolescents and public concerns of negative psycho-social outcomes, Snapchat streaks and the associated variables have received limited empirical attention [22]. The purpose of this study is to address this gap in current literature by examining the relationship between participation in Snapchat streaks with peers and problematic smartphone use, fear of missing out, and smartphone self-control among early adolescents. Our study was guided by the following research question:

> **RQ1:** What is the relationship between participation in Snapchat streaks and problematic smartphone use, fear of missing out, and smartphone self-control among early adolescents?

Early adolescents are a specifically relevant population for this study as friendships hold significant importance in their development [46]. Snapchat streaks could be an important way to communicate within their friendships for some early adolescents [22]. Especially, early adolescents may lack the maturity of their older counterparts to tackle the challenges associated with engaging in Snapchat streaks. These streaks are governed by specific social norms, which may add to pressure within friendships to be constantly online and accessible [22,47,48]. The findings of our exploratory study could contribute to the field by offering insights into the traits of young individuals who participate in Snapchat streaks, thus contributing to both theoretical understanding and practical application.

Methods

Sample and procedures

The data were collected using paper-and-pencil surveys in the Spring of 2019 as part of Belgian Early Adolescent Risk Study, which was conducted in 12 schools in the Dutch-speaking area of Belgium. The surveys were administered during class time among the first three years of secondary education. Within each of the 14 schools that agreed to participate, the administration chose class groups and invited every student in those groups to take part in the survey. The surveys were conducted during class hours in the first three years of secondary education, which corresponds to middle school students and high school freshmen. Passive parental consent and consent from the schools' principals were obtained prior to the study. IRB approval was granted by the second author's institution.

Participants were between 12 and 16 years old. Given the scope of our study, in which we focus on early adolescents, we have removed adolescents who were over 15 years old from our analysis ($n = 77$). Additionally, participants who did not answer all questions used for the items measured, or were inconsistent in their answers, indicating inattentive reading (e.g., indicated they did not use Snapchat, but reported on their current Snapchat streak), were excluded ($n = 171$). The final sample consisted of 2483 adolescents. The age of the participants in the study ranged from 12 to 15 years old, with a mean age of 13.46 years ($SD = 0.894$). A total of 1388 girls (55.9%) and 1046 boys (42.1%) participated, 49 (2.0%) did not indicate their gender.

Measures

Demographic variables: Participants were asked to indicate their age and gender (male/female).

Snapchat use: Participants were first asked whether they used Snapchat (yes/no). The descriptive information for all the variables is included in Table 1.

Involvement in a Snapchat streak: Subsequently, participants who indicated that they used Snapchat were asked: "do you currently have a Snapchat streak with one or more people" (0 = *no streak*, 1 = *current streak*). A Snapchat streak was defined as "you and a friend send each other during three consecutive days a snap (not a chat), each time within 24 hours. When this happens, you'll get a flame next to their name".

Problematic smartphone use: Participants' problematic smartphone use was assessed using a 4-item scale, which asked them to rate their frequency of engagement on a 5-point scale. The scale ranged from 0, indicating "hardly ever," to 4, indicating "very often". The items included: "While I am with my partner, family or friends, I am very much occupied with my smartphone.", "While I am with my partner, family or friends, I am unable to use my smartphone less", "While I am with my partner, family or friends, I find it hard to stay focused because I am busy with my smartphone" and "While I am with my partner, family or friends I ignore them sometimes because I am on my smartphone". The scale was reliable ($\alpha = 0.72$).

Fear of Missing Out (FOMO): To measure participants FOMO, we used four items on a five point scale (1 = never, 5 = very often) that were adapted from the Fear of Missing Out measure by [49], such as "I am restless/nervous when I forgot my phone and cannot check my social media to see what my friends are doing", or "I am restless when I feel I am missing important updates when I don't have access to social media". The scale was reliable ($\alpha = 0.81$).

Table 1
Descriptive statistics.

	N	Mean	SD	Min	Max	No. of items	α
Age	2474	13.46	0.894	12	15		
Gender	2434	0.57	0.495	0	1		
Snapchat use	2425	0.79	0.404	0	1		
Snapchat streak	1927	0.77	0.420	0	1		
Number of people	1488	28.45	32.083	1	350		
Longest streak	1477	209.45	193.077	1	1024		
Problematic smartphone use	2483	0.7940	0.62503	0	3.75	4	0.72
FOMO	2483	1.7625	0.75201	1	5	4	0.81
Social Media Self Control	2483	2.2721	0.86215	1	5	3	0.79

Social Media Self Control: Lastly, social media self-control was measured using an adapted version of the "brief measure of social media self-control failure" [43]. A prompt stated "I use social media or my smartphone even though…" after which respondents were presented with three items on a 5-point scale (1 = hardly ever, 5 = very often). The scale included statements such as "…I need to postpone other things" and "…it causes me to do other activities less well". The scale was reliable ($\alpha = 0.79$).

Data analysis

The data were analyzed using SPSS V.26.0 (IBM, Armonk, NY). For the descriptive statistics, the differences between gender and involvement in Snapchat streaks were examined using a chi-square test. The gendered engagement in Snapchat streaks, and the number of people adolescents engaged in Snapchat streaks was analyzed using a *t*-test. The age in relation to the number of people and number of days adolescents maintained Snapchat streaks was analyzed using an ANOVA. Lastly, the engagement in Snapchat streaks and the relationship with problematic smartphone use, FOMO, and social media self-control was measured using a *t*-test, and problematic smartphone use, FOMO, and social media self-control and their relationship with the number of people and days Snapchat streaks were maintained with was measured using correlation.

Results

Engagement in Snapchat streaks

Girls who were on Snapchat were found to be significantly more likely to engage in Snapchat streaks than boys (see Table 2). In our study 83.3% of the girls in our sample had a current Snapchat streak versus 66.9% of the boys in our sample.

Age was considered as a factor influencing engagement in maintaining Snapchat streaks (see Table 2). The results indicate no significant differences for the relationship between age and the engagement in maintaining Snapchat streaks.

A *t*-test comparing people who do engage in Snapchat streaks and those who do *not* engage in Snapchat streaks, shows that adolescents engaging in Snapchat streaks show statistical significance for an increase in problematic smartphone use ($t(1925) = 6.22$, $p < 0.001$). FOMO and social media self-control were not significantly related to engagement in Snapchat streaks (see Table 3).

Number of people

When considering the number of people participants had Snapchat streaks with, no significant difference was noted between girls and boys ($t(729.40) = -1.87$, $p = 0.68$). Age did show a statistical significance in the difference between the age groups. ($F(3, 1513) = [4.45]$, $p < 0.005$), however the effect size was small at $\eta 2 = 0.01$.

When correlating the number of people adolescents maintain Snapchat streaks with problematic smartphone use, FOMO, and social media self-control, each variable showed a significant relationship (see Table 4). There was a positive relationship between number of people they maintain streaks with and problematic smartphone use ($r = 0.123$, $n = 1488$, $p < 0.001$), implying that with more problematic smartphone use adolescents were also more likely to have more people they maintained streaks with. FOMO also showed a positive relationship ($r = 0.146$, $n = 1488$, $p < 0.001$), implying that with more FOMO, adolescents also had more people they maintained streaks with. Lastly, social media self-control showed a positive relationship as well ($r = 0.106$, $n = 1488$, $p < 0.001$), implying that adolescents with less self-control had maintained more streaks.

Number of days

When obtaining an insight in the number of days that adolescents maintained streaks for, gender and age, both showed a statistically significant relationship. For gender, girls ($M = 222.94$, $SD = 196.93$) on average had longer Snapchat streaks than boys ($M = 179.25$, $SD = 182.89$; $t(938.58) = -4.01$, $p < 0.001$. Age also significantly influenced the length of the streaks ($F(3, 1466) = [20.272]$, $p < 0.001$), with older adolescents having longer streaks, however, the effect size was small at $\eta 2 = 0.04$.

In correlation with the length of the longest streak, problematic smartphone use, FOMO, and social media self-control all showed statistically significant results (see Table 4). Problematic smartphone use is related to longer streaks ($r = 0.115$, $n = 1477$, $p < 0.001$). For FOMO, ($r = 0.175$, $n = 1477$, $p < 0.001$), the more FOMO adolescents experience, the longer their streaks. Lastly, when adolescents have less social media self-control ($r = 0.087$, $n = 1477$, $p < 0.001$), they are more likely to also have longer streaks.

Discussion

As parents, the media, and some politicians voiced concerns about the engagement of adolescents on Snapchat, we wanted to examine how FOMO, problematic smartphone use, and social media self-control impact adolescent use of Snapchat with peers. Our study aimed to generate a better understanding of Snapchat streak usage and its associations with forms of problematic social media use. Our study found a significant difference between girls and boys and the maintenance of their Snapchat streaks, whereby girls were more likely to engage in Snapchat streaks with peers and for longer periods of time, but not with significantly more people. The gender difference may be explained by the fact that girls have been found to be more likely to engage in forms of online self-disclosure, as a way to connect with others and maintain social connections, and to establish bonding social capital through their online communication (Van Gool et al., 2015). Adolescents' age did significantly influence not only the engagement in maintaining Snapchat streaks, but also the number of people and the length of time they maintained the streaks for.

Interestingly, only problematic smartphone use significantly influenced students' engagement (yes or no) in Snapchat streaks. Indicating that students who engage in problematic smartphone use more likely maintain daily, consistent, interaction with one person. However, once maintaining a streak, the number of people and the maximum number of days adolescents engage in Snapchat streak are all significantly correlated with FOMO, problematic smartphone use, and social media self-control. Although speculative, this may potentially be explained by adolescents' fear of not feeling connected to others through their smartphone could drive a feeling of losing out their social interactions if they do not maintain a Snapchat streak.

The effect size of all variables, though statistically significant, are relatively small. It can therefore be said that, unless engaging in Snapchat streaks may be risky in other contexts (e.g., adolescents engaging in password sharing [50], or other forms of social media addiction [51], Snapchat streaks themselves are not typically related to strong changes. The engagement in Snapchat streaks might, in fact, be part of a normative communication process of the adolescent population not

Table 2
Snapchat streaks and demographic characteristics of the participants.

Variable		Snapchat Streak		Snapchat Streak People		Snapchat Streak Days	
		Yes	No	Mean	SD	Mean	SD
Gender	Boys	66.8%	33.2%	25.93	37.009	179.25	182.895
	Girls	83.0%	17.0%	29.32	39.387	222.94	196.929
Age	12	73.9%	26.1%	19.09	21.326	134.08	158.676
	13	76.5%	23.5%	28.73	35.286	184.91	179.001
	14	78.3%	21.7%	30.49	32.140	237.64	197.797
	15	76.9%	23.1%	29.64	28.821	259.80	214.393

Table 3
Snapchat streaks and main study variables.

Variable	Snapchat Streak Yes (n = 1488)		Snapchat Streak No (n = 439)		t(1925)	p	Cohen's d
	M	SD	M	SD			
Problematic smartphone use	0.854	.616	.786	.642	6.222	<0.001	.622
FOMO	1.905	.778	1.646	.668	2.013	.443	.754
Social Media Self Control	2.371	.855	2.251	.876	2.562	.505	.856

Table 4
Bivariate correlations between variables.

	1	2	3	4	5
1. Number of people	1.000	.485**	.123**	.146**	.106**
2. Number of days	.485**	1.000	.225**	.175**	.087**
3. Problematic smartphone use	.123**	.115**	1.000	.440**	.392**
4. FOMO	.146**	.175**	.440**	1.000	.404**
5. Social Media Self Control	.106**	.087**	.392**	.404**	1.000

** $p < 0.001$.

necessarily driven, in strong degrees, by FOMO, problematic smartphone use, and social media self-control. Our study does contribute to cumulative evidence that these novel forms of interpersonal communication do not necessarily have to be understood from a risk perspective and that the public concern about these gamified forms of interpersonal communication may be overstated [26].

In practice, this might mean for parents and educators that they should not be too concerned about children's engagement with Snapchat streaks as it pertains to their children's FOMO, problematic smartphone use, and social media self-control. Although each of these items have a statistically significant correlation with the number of people and the number of days adolescents maintained Snapchat streaks, it is important to acknowledge that there may be other factors not addressed in this study that could have a more significant impact on adolescents' use of Snapchat streaks. Unless adolescents are engaging in potentially harmful interactions on Snapchat (e.g. interactions with strangers, interactions with adults, adult sexual solicitation, bullying), adolescents' use of Snapchat with other peers may not necessarily be related to negative digital media use.

Limitations and suggestions for future research

Some limitations in this research need to be kept in mind when interpreting the results of our study. First, the cross-sectional data were self-reported by a convenience sample of adolescents. To collect the data for our study, we relied on a convenience sample of adolescents who self-reported their information. It is worth noting that this type of cross-sectional data collection approach may have limitations in terms of its generalizability to larger populations. Our study was conducted in Belgium, additional cross-cultural and cross-national research is warranted. Despite efforts to specify the specific time frames and reference frames of adolescents' Snapchat use, the respondents may inaccurately recall their Snapchat use. Future work could use alternative data collection strategies and alternative sampling techniques that can help to reduce bias and longitudinal designs are needed to track behaviors over time. Our study was only able to collect binary-gender data of the young adolescents. Future work should use alternative gender measures that allow to capture a broad range of gender identities. Additionally, future work should seek to understand how problematic smartphone use, FOMO, and social media self-control compares to other social media platforms using similar principles, such as the upcoming platform BeReal.

Our exploratory study raises additional questions for future theory-driven research to gain a deeper understanding of the antecedents and consequences of engagement in Snapchat streaks. For example, Uses and Gratifications Theory [52] could provide a useful framework to understand users' motivations to maintain Snapchat streaks and to understand the gamified nature of the communication. The Differential Susceptibility to Media Effects Model suggests that media use has differential antecedents and consequences based on dispositional, developmental, and social susceptibility variables [53]. Future work could investigate whether particular personality or developmental factors are linked to the use of Snapchat streaks, and whether they lead to different usage patterns and outcomes. It is conceivable that perhaps particular attachment styles and certain friendship contexts could be related with a more preoccupied use of Snapchat streaks, which could be explored by future work. Additionally, our study did not examine the broader social context in which these gamified interactions occur. A Snapchat streak is both an intense social ritual as well as a gamified form of interpersonal interaction that comes with social rewards. The role of peer norms and peer pressure in engaging and maintaining Snapchat streaks could be examined using theories that focus on social norms, such as the Prototype Willingness Model [54] and Social Learning Theory [55].

Future work could also examine the potential consequences of a sustained or failed Snapchat streak through the lens of interpersonal communication theories, such as Expectancy Violations Theory [56]. Expectancy Violations Theory has been previously used to understand other digital friendship dynamics, such as unfriending behaviors on Facebook [57] , and could be applied to understand the consequences of a failed Snapchat streak on friendships. Health-focused studies could focus on how Snapchat streaks are associated with certain (mental) health negative outcomes through the lens of Problem Behavior Theory [58].

The role of Snapchat streaks in friendships and the maintenance of adolescent friendships has been documented by several researchers (e.g., [22,21]). Yet, more work should be conducted on the long-term effects of image-based social media applications that center around ephemeral content. For instance, past research has discovered that Snapchat streaks can contribute to the digital stress experienced by adolescents, which stems from numerous notifications and the peer pressure to preserve the streak [22,24]. Future research could investigate both the potential negative and positive mental health implications of engaging in Snapchat streaks. This includes examining the stress and anxiety that may result from the persistent pressure to be online, as well as exploring how Snapchat streaks might promote a sense of belonging and strengthen friendships [59].

Additionally, future work could explore how adults perceive gamified interactions, such as Snapchat streaks, within adult relationships. The motivations for media use and the associated outcomes of Snapchat differ greatly between adolescents and adults and are difficult to compare. Given that adult relationships tend to be more stable than adolescent ones, it would be interesting to explore the significance of such interactions in the context of adult relationships.

To gain a deeper understanding of the impact of Snapchat streaks on adolescent relationships, research could investigate how they are associated with adolescents' perceptions of social capital and the quality of their friendships. Indeed, as streaks reveal the frequency of interactions, adolescents can use them to see who their friends are interacting with the most. This visibility may result in feelings of friendship jealousy or annoyance within the peer group. Alternatively, it could also increase a sense of connection among friends. Therefore, exploring the contribution of Snapchat streaks to adolescent relationship maintenance could provide insight on their role in maintaining friendships.

Conclusion

This study aimed to provide insight into the correlates of Snapchat streaks, an emerging, ritualized communication behavior among adolescents. In our sample, girls were more likely than boys to engage in Snapchat streaks and the average number of days that a streak was held. An age difference was only observed for the number of people and the number of days they maintained snapchat streaks with. Problematic smartphone use was related to the engagement in Snapchat streaks. Lastly, FOMO, problematic smartphone use, and social media self-control were correlated with the number of people and the number of days adolescents maintained Snapchat streaks with friends, albeit it being a weak relationship.

Funding

The study was supported by a postdoctoral research grant of the Research Foundation – Flanders under grant number 12J8719N and a grant by the Research Fund of the University of Antwerp under grant number FFB190303. The study design, data collection, analysis, interpretation of the data, writing of the report and the decision to submit the article for publication were the sole responsibility of the authors and were in no way influenced by the funding institutions.

Financial disclosure

The authors have no financial relationships relevant to this article to disclose.

Declaration of Competing Interest

The authors have no conflict of interest to disclose.

Data availability

The data that has been used is confidential.

References

[1] H. Vanwesenbeeck, [illegible], [illegible], [illegible], [illegible], R. De Wolf, J. Van Ouytsel, K. Van Damme, J. Vanderberg, [illegible], [illegible], [illegible], S. De Pauw, [illegible], [illegible], Onderzoeksrapport Apestaartjaren: de digitale leefwereld van kinderen en jongeren [Research report Apestaartjaren: The digital environment of children and youth], Mediawijs (2022).
[2] S. Handyside, J. Ringrose, Snapchat memory and youth digital sexual cultures: mediated temporality, duration and affect, J. Gend. Stud. 26 (3) (2017) 347–360, https://doi.org/10.1080/09589236.2017.1280384.
[3] G. Mittmann, K. Woodcock, S. Dörfler, I. Krammer, I. Pollak, B. Schrank, "TikTok is my life and Snapchat is my ventricle": a mixed-methods study on the role of online communication tools for friendships in early adolescents, J. Early Adolesc. 42 (2) (2022) 172–203, https://doi.org/10.1177/02724316211020368.
[4] L. Piwek, A. Joinson, "What do they snapchat about?" Patterns of use in time-limited instant messaging service, Comput. Hum. Behav. 54 (2016) 358–367, https://doi.org/10.1016/j.chb.2015.08.026.
[5] M.A. Throuvala, M.D. Griffiths, M. Rennoldson, D.J. Kuss, Motivational processes and dysfunctional mechanisms of social media use among adolescents: a qualitative focus group study, Comput. Hum. Behav. 93 (2019) 164–175, https://doi.org/10.1016/j.chb.2018.12.012.
[6] J.B. Bayer, N.B. Ellison, S.Y. Schoenebeck, E.B. Falk, Sharing the small moments: ephemeral social interaction on Snapchat, Inf. Commun. Soc. 19 (7) (2016) 956–977, https://doi.org/10.1080/1369118X.2015.1084349.
[7] J. Kofoed, M.C. Larsen, A snap of intimacy: photo-sharing practices among young people on social media, First Monday 21 (11) (2016), https://doi.org/10.5210/fm.v21i11.6905.
[8] A. Vermeulen, H. Vandebosch, W. Heirman, #Smiling, #venting, or both? Adolescents' social sharing of emotions on social media, Comput. Hum. Behav. 84 (2018) 211–219, https://doi.org/10.1016/j.chb.2018.02.022.
[9] S. Alhabash, M. Ma, A tale of four platforms: motivations and uses of Facebook, Twitter, Instagram, and Snapchat among college students? Soc. Media + Soc. 3 (1) (2017) 1–13, https://doi.org/10.1177/2056305117691544.
[10] E.P. Abell, K. Tyson, K. Morefield, SnapChat this, Instagram that: the interplay of motives and privacy affordances in college students' sharing of food porn, Telematics Inf. 74 (2022), 101889, https://doi.org/10.1016/j.tele.2022.101889.
[11] D.S. Strassberg, J.K. Rullo, J.E. Mackaronis, The sending and receiving of sexually explicit cell phone photos ("Sexting") while in high school: one college's students' retrospective reports, Comput. Hum. Behav. 41 (0) (2014) 177–183, https://doi.org/10.1016/j.chb.2014.09.008.
[12] J.R. Temple, H. Choi, Longitudinal association between teen sexting and sexual behavior, Pediatrics 134 (5) (2014) e1287–e1292, https://doi.org/10.1542/peds.2014-1974.
[13] S. De Ridder, Sexting as sexual stigma: the paradox of sexual self-representation in digital youth cultures, Eur. J. Cultural Stud. (2018), https://doi.org/10.1177/1367549418810080, 1367549418810080.
[14] J. Van Ouytsel, E. Van Gool, M. Walrave, K. Ponnet, E. Peeters, Sexting: adolescents' perceptions of the applications used for, motives for, and consequences of sexting, J. Youth Stud. 20 (4) (2017) 446–470, https://doi.org/10.1080/13676261.2016.1241865.
[15] M.B.R. Mandau, Snaps", "screenshots", and self-blame: a qualitative study of image-based sexual abuse victimization among adolescent Danish girls, J. Child. Media 15 (3) (2021) 431–447, https://doi.org/10.1080/17482798.2020.1848892.
[16] J.R. Moran, K.J. Salerno, T.J. Wade, Snapchat as a new tool for sexual access: are there sex differences? Pers. Individ. Dif. 129 (2018) 12–16, https://doi.org/10.1016/j.paid.2018.02.040.
[17] J. Sanderson, M.R. Weathers, Snapchat and child sexual abuse in sport: protecting child athletes in the social media age, Sport Manag. Rev. 23 (1) (2020) 81–94, https://doi.org/10.1016/j.smr.2019.04.006.
[18] D. Hristova, J. Dumit, A. Lieberoth, T. Slunecko, Snapchat streaks: how adolescents metagame gamification in social media, PsyArXiv Preprints (2019), https://doi.org/10.31234/osf.io/[illegible].
[19] D. Hristova, S. Jovicic, B. Göbl, S. de Freitas, T. Slunecko, Why did we lose our snapchat streak?". Social media gamification and metacommunication, Comput. Hum. Behav. Rep. 5 (2022) (2022), https://doi.org/10.1016/j.chbr.2022.100172.
[20] S. McGlotten, Streaking, TDR/The Drama Rev. 63 (4) (2019) 152–171, https://doi.org/10.1162/dram_a_00881.
[21] D. Hristova, A. Lieberoth, How socially sustainable is social media gamification? A look into Snapchat, Facebook, Twitter and Instagram, in: A. Spanellis, J.T. Harviainen (Eds.), Transforming Society and Organizations through Gamification: From the Sustainable Development Goals to Inclusive Workplaces, Springer International Publishing, 2021, pp. 225–245, https://doi.org/10.1007/978-3-030-68207-1_12.
[22] D. De Groote, J. Van Ouytsel, Digital stress within early adolescents' friendships–A focus group study from Belgium, Telematics Inf. 73 (2022), 101877, https://doi.org/10.1016/j.tele.2022.101877.
[23] J. Van Ouytsel, D. De Groote, Research brief: early adolescents' perceptions of the motivations and consequences of sharing passwords with friends in Belgium, J. Child Media (2022) 1–12, https://doi.org/10.1080/17482798.2022.[illegible].
[24] D.L. Weber, A.S.M. Bygholm, A.M. Kanstrup, Digital coping strategies of young people living with cognitive disabilities: using emotion-focused and problem-focused coping in digital relationships, Disabil. Soc. (2023) 1–25, https://doi.org/10.1080/09687599.2023.2181767.
[25] Snapchat. (n.d.). Friend emojis. https://support.snapchat.com/en-US/a/friend-emojis.
[26] S. Utz, N. Muscanell, C. Khalid, Snapchat elicits more jealousy than Facebook: a comparison of Snapchat and Facebook use, Cyberpsychol. Behav. Soc. Netw. 18 (3) (2015) 141–146, https://doi.org/10.1089/cyber.2014.0479.
[27] M.J. Vaterlaus, K. Barnett, C. Roche, J.A. Young, "Snapchat is more personal": an exploratory study on Snapchat behaviors and young adult interpersonal relationships, Comput. Hum. Behav. 62 (2016) (2016) 594–601, https://doi.org/10.1016/j.chb.2016.04.029.
[28] B. Korkmazer, S. De Ridder, S. Van Bauwel, Reporting on young people, sexuality, and social media: A discourse theoretical analysis, J. Youth Stud. 23 (3) (2020) 323–339, https://doi.org/10.1080/13676261.2019.1603365.
[29] M. Murphy, Streaks Game Fears Snapchat Streaks 'should be stopped' Claim Worried Parents Who Fear It Puts Teens At Risk, The Sun, 2017. Retrieved from, https://www.thesun.co.uk/tech/3477571/snapchat-streaks-should-be-banned-claim-worried-parents-who-fear-it-puts-teens-at-risk/.
[30] 116th Congress - the Senate of the United States. Social media addiction reduction technology act or the SMART Act, 2314, 116th Congress - the Senate of the United States (2019).
[31] Hawley, J. (2019). Sen. Hawley introduces legislation to curb social media addiction. Retrieved from https://www.hawley.senate.gov/sen-hawley-introduces-legislation-curb-social-media-addiction.
[32] B. Fintech, Oh Snap: U.S. Senator Proposes Bill to Ban Snapchat 'Snapstreaks', Reuters, 2019. Retrieved from, https://www.reuters.com/article/us-usa-senate-social-media/oh-snap-u-s-senator-proposes-bill-to-ban-snapchat-snapstreaks-idUSKCN1UN2GA.
[33] N. Bloemen, D. De Coninck, Social media and fear of missing out in adolescents: the role of family characteristics, Soc. Media + Soc. 6 (4) (2020), 2056305120965517, https://doi.org/10.1177/2056305120965517.
[34] I. Beyens, E. Frison, S. Eggermont, I don't want to miss a thing": adolescents' fear of missing out and its relationship to adolescents' social needs, Facebook use, and Facebook related stress, Comput. Hum. Behav. 64 (2016) 1–8, https://doi.org/10.1016/j.chb.2016.05.083.
[35] V. Franchina, M. Vanden Abeele, Van Rooij, G. Lo Coco, L. De Marez, Fear of missing out as a predictor of problematic social media use and phubbing behavior among Flemish adolescents, Int. J. Environ. Res. Public Health 15 (10) (2018) 2319, https://doi.org/10.3390/ijerph15102319.

[36] I. Bartau-Rojas, A. Aierbe-Barandiaran, E. Oregui-González, Parental mediation of the internet use of primary students: beliefs, strategies and difficulties, Media Educ. Res. J. 26 (1) (2018) 2–6, https://doi.org/10.3916/C54-2018-07.

[37] S. Coskun, G. Karayagız Muslu, Investigation of problematic mobile phones use and fear of missing out (FoMO) level in adolescents, Commun. Ment. Health J. 55 (6) (2019) 1004–1014, https://doi.org/10.1007/s10597-019-00422-8.

[38] M. Kwon, J.Y. Lee, W.Y. Won, J.W. Park, J.A. Min, C. Hahn, X. Gu, J.-H. Choi, D.J. Kim, Development and validation of a smartphone addiction scale (SAS), PLoS One 8 (2) (2013) e56936, https://doi.org/10.1371/journal.pone.0056936.

[39] T. Panova, X. Carbonell, Is smartphone addiction really an addiction? J. Behav. Addict. 7 (2) (2018) 252–259, https://doi.org/10.1556/2006.7.2018.49.

[40] S.H. Liu, M.C. Yin, T.H. Huang, Adolescents' interpersonal relationships with friends, parents, and teachers when using Facebook for interaction, Creative Educ. 4 (5) (2013) 335–339, https://doi.org/10.4236/ce.2013.46048.

[41] G.E. Maynard, E.G.A. Abraham, The relationship between use of technology and parent-adolescent social relationship, J. Educ. Practice 7 (14) (2016) 168–179.

[42] M.J. George, C.L. Odgers, Seven fears and the science of how mobile technologies may be influencing adolescents in the digital age, Perspect. Psychol. Sci. 10 (6) (2015) 832–851, https://doi.org/10.1177/1745691615596788.

[43] J. Du, G.M. van Koningsbruggen, F. Kerkhof, A brief measure of social media self-control failure, Comput. Hum. Behav. 84 (2018) 68–75, https://doi.org/10.1016/j.chb.2018.02.002.

[44] S.S. Abrams, S. Walsh, Gamified vocabulary: online resources and enriched language learning, J. Adolesc. Adult Literacy 58 (1) (2014) 49–58, https://doi.org/10.1002/jaal.315.

[45] Y. Yoshida-Montezuma, M. Ahmed, O. Ezezika, Does gamification improve fruit and vegetable intake in adolescents? A systematic review, Nutr. Health 26 (4) (2020) 347–366, https://doi.org/10.1177/0260106020936143.

[46] A. Bowker, Predicting friendship stability during early adolescence, J. Early Adolesc. 24 (2) (2004) 85–112, https://doi.org/10.1177/0272431603262666.

[47] G. Daenekindt, K. Verswijvel, L. De Marez, I'll be online for you: the role of social media in the achievement of adolescents' social goals, In: M.H. Jones (Ed.), Social Goals in the classroom: Findings On Student Motivation and Peer Relations, Routledge, 2019.

[48] R.G. Steele, J.A. Hall, J.L. Christofferson, Conceptualizing digital stress in adolescents and young adults: toward the development of an empirically based model, Clin. Child Fam. Psychol. Rev. 23 (1) (2020) 15–26, https://doi.org/10.1007/s10567-019-00300-5.

[49] A.K. Przybylski, K. Murayama, C.R. DeHaan, V. Gladwell, Motivational, emotional, and behavioral correlates of fear of missing out, Comput. Hum. Behav. 29 (4) (2013) 1841–1848, https://doi.org/10.1016/j.chb.2013.02.014.

[50] V. Povilaitis, Smartphone-free summer camp: adolescent perspectives of a leisure context for social and emotional learning, World Leis J. 61 (4) (2019) 276–290, https://doi.org/10.1080/16078055.2019.1661104.

[51] R. Noë, L.D. Turner, D.E.J. Linden, S.M. Allen, B. Winkens, R.M. Whitaker, Identifying indicators of smartphone addiction through user-app interaction, Comput. Hum. Behav. 99 (2019) (2019) 56–65, https://doi.org/10.1016/j.chb.2019.04.023.

[52] T.E. Ruggiero, Uses and gratifications theory in the 21st century, Mass Commun. Soc. 3 (1) (2000) 3–37, https://doi.org/10.1207/S15327825MCS0301_02.

[53] P.M. Valkenburg, J. Peter, The differential susceptibility to media effects model, J. Commun. 63 (2) (2013) 221–243, https://doi.org/10.1111/jcom.12024.

[54] F.X. Gibbons, M. Gerrard, H. Blanton, D.W. Russell, Reasoned action and social reaction: willingness and intention as independent predictors of health risk, J. Pers. Soc. Psychol. 74 (1998) 1164–1180, https://doi.org/10.1037/0022-3514.74.5.1164.

[55] R.L. Akers, M.D. Krohn, L. Lanza-Kaduce, M. Radosevich, Social learning and deviant behavior: a specific test of a general theory, Am. Sociol. Rev. 44 (4) (1979) 636–655, https://doi.org/10.2307/2094592.

[56] Burgoon, J.K. (2015). Expectancy violations theory. The International Encyclopedia of Interpersonal Communication, 1–9.

[57] J.L. Bevan, P.C. Ang, J.B. Fearns, Being unfriended on Facebook: an application of expectancy violation theory, Comput. Hum. Behav. 33 (2014) 171–178, https://doi.org/10.1016/j.chb.2014.01.029.

[58] R. Jessor, The origins and development of problem behavior theory, Adv. Responsible Adolesc. Dev. (2016).

[59] H.L. Storer, C.F. Scott, M. Rodriguez, E.S. Nyerges, Technology is a "blessing and a curse": the perceived risks and benefits of digital technology adoption at domestic violence organizations that serve teens, J. Technol. Hum. Serv. 41 (1) (2023) 96–124, https://doi.org/10.1080/15228835.2023.2179158.

5

THIS ARTICLE HAS BEEN CORRECTED. SEE LAST PAGE.

Psychology of Popular Media

ISSN: 2689-6567

2025, Vol. 14, No. 1, 1–11
https://doi.org/10.1037/ppm0000536

Limiting Social Media Use Decreases Depression, Anxiety, and Fear of Missing Out in Youth With Emotional Distress: A Randomized Controlled Trial

Christopher G. Davis[1] and Gary S. Goldfield[2]
[1] Department of Psychology, Carleton University
[2] Healthy Active Living and Obesity Research Group, Children's Hospital of Eastern Ontario Research Institute, Ottawa, Canada

Reports demonstrating modest but significant correlations between heavy social media use (SMU) and poorer
mental health in youth have led many to
harmed by heavy SMU, distressed yout
internally examine the effects of reduci
missing out (FoMO), and sleep in yout
assign 220 youth aged 17–25 years to
asked to reduce smartphone-based SMU
tions. SMU was objectively measured
subjectively assessed at baseline and fol
the intervention group showed significa
and greater increases in sleep. No effec
imately 1 hr/day may be a feasible, inex
toms of depression, anxiety, and FoMO

fomo

Public Policy Relevance Statement
A brief 4-week intervention using
heavy social media users reporting
in symptoms of depression, anxiety
a comparable control group. Reduc
among a vulnerable population of

Keywords: social media, depression,

Adolescence and young adulthood spanning ages 17–25 years is a period of development that is often referred to as youth and is widely considered to be among the most vulnerable periods for the development of mental illness due to the unique social, physical, emotional, and neurobiological changes (Paus et al., 2008). Approximately 20% of youth will be diagnosed with a mental disorder in any given year (Comeau et al., 2019). Depression and anxiety disorders are the most common forms of mental illness with recent epidemiological data (collected during the COVID-19 pandemic) indicating that 36% and 23% of Canadian youth (18–24 years) met clinical thresholds for major depressive disorder and generalized anxiety disorder (GAD), respectively (Statistics Canada, 2021).

Paralleling the growing prevalence of anxiety and depressive disorders among youth has been the growing use of social media as a primary form of social interaction, with approximately 81.3% of Canadian youth reporting moderate-to-heavy (2 hr or more) daily use (Sampasa-Kanyinga et al., 2019), and more than 96% using at least one social media platform (Sampasa-Kanyinga et al., 2019; Statistics Canada, 2019). Moreover, data suggest that the vast majority of social media use (SMU) in youth occurs on smartphones; approximately 95% of American teens report daily access to a smartphone, while 45% reported being online almost constantly (Anderson & Jiang, 2018). Although SMU is appealing in that it provides opportunities for social connection and efficient communication, constant access to these platforms combined with the high volume of notifications that effectively tie youth to their smartphones has sparked concerns about the psychological consequences of heavy SMU among youth.

Several narrative and meta-analytic reviews have indicated small but statistically significant correlations ($r = .11$–$r = .17$) between duration of SMU and depressive symptoms (Orben, 2020). Similarly, a systematic review of 13 studies in youth found that time spent

This article was published Online First April 22, 2024.
Christopher G. Davis https://orcid.org/0000-0003-3851-080X
Gary S. Goldfield https://orcid.org/0000-0001-6216-7834
The authors report that there are no competing interests to declare. The data that support the findings of this study are available in the Open Science Framework at https://osf.io/5c847/.

Christopher G. Davis served as lead for data curation, formal analysis, project administration, and visualization. Gary S. Goldfield served as lead for supervision and validation. Christopher G. Davis and Gary S. Goldfield contributed equally to conceptualization, writing–original draft, writing–review and editing, and methodology.
Correspondence concerning this article should be addressed to Gary S. Goldfield, Healthy Active Living and Obesity Research Group, Children's Hospital of Eastern Ontario Research Institute, 401 Smyth Road, Ottawa, ON K1H 8L1, Canada. Email: ggoldfield@cheo.on.ca

THIS ARTICLE HAS BEEN CORRECTED. SEE LAST PAGE.

American Psychological Association

Psychology of Popular Media

ISSN: 2689-6567

2025, Vol. 14, No. 1, 1–11
https://doi.org/10.1037/ppm0000536

Limiting Social Media Use Decreases Depression, Anxiety, and Fear of Missing Out in Youth With Emotional Distress: A Randomized Controlled Trial

Christopher G. Davis[1] and Gary S. Goldfield[2]

[1] Department of Psychology, Carleton University

[2] Healthy Active Living and Obesity Research Group, Children's Hospital of Eastern Ontario Research Institute, Ottawa, Canada

Reports demonstrating modest but significant correlations between heavy social media use (SMU) and poorer mental health in youth have led many to suggest that heavy SMU is culpable. Although many youth may not be harmed by heavy SMU, distressed youth may be particularly vulnerable. The aim of this study was to experimentally examine the effects of reducing SMU on smartphones on symptoms of depression, anxiety, fear of missing out (FoMO), and sleep in youth with emotional distress. A randomized controlled trial was used to assign 220 youth aged 17–25 years to either an intervention or control group. The intervention group was asked to reduce smartphone-based SMU to 1 hr/day for 3 weeks while the control group had no SMU restrictions. SMU was objectively measured daily via tracking systems in smartphones. Mental health and sleep were subjectively assessed at baseline and following the 3-week intervention period. Compared to the control group, the intervention group showed significantly greater reductions in symptoms of depression, anxiety, and FoMO, and greater increases in sleep. No effects of gender were detected. Reducing SMU on smartphones to approximately 1 hr/day may be a feasible, inexpensive, and effective method of increasing sleep and reducing symptoms of depression, anxiety, and FoMO among distressed youth.

Public Policy Relevance Statement

A brief 4-week intervention using screen time trackers showed that reducing social media use among heavy social media users reporting symptoms of depression or anxiety yielded significant reductions in symptoms of depression, anxiety, and fear of missing out, and increased hours of sleep relative to a comparable control group. Reducing social media use may be a feasible method of reducing distress among a vulnerable population of heavy social media users.

Keywords: social media, depression, anxiety, fear of missing out, social networking sites

Adolescence and young adulthood spanning ages 17–25 years is a period of development that is often referred to as youth and is widely considered to be among the most vulnerable periods for the development of mental illness due to the unique social, physical, emotional, and neurobiological changes (Paus et al., 2008). Approximately 20% of youth will be diagnosed with a mental disorder in any given year (Comeau et al., 2019). Depression and anxiety disorders are the most common forms of mental illness with recent epidemiological data (collected during the COVID-19 pandemic) indicating that 36% and 23% of Canadian youth (18–24 years) met clinical thresholds for major depressive disorder and generalized anxiety disorder (GAD), respectively (Statistics Canada, 2021).

Paralleling the growing prevalence of anxiety and depressive disorders among youth has been the growing use of social media as a primary form of social interaction, with approximately 81.3% of Canadian youth reporting moderate-to-heavy (2 hr or more) daily use (Sampasa-Kanyinga et al., 2019), and more than 96% using at least one social media platform (Sampasa-Kanyinga et al., 2019; Statistics Canada, 2019). Moreover, data suggest that the vast majority of social media use (SMU) in youth occurs on smartphones; approximately 95% of American teens report daily access to a smartphone, while 45% reported being online almost constantly (Anderson & Jiang, 2018). Although SMU is appealing in that it provides opportunities for social connection and efficient communication, constant access to these platforms combined with the high volume of notifications that effectively tie youth to their smartphones has sparked concerns about the psychological consequences of heavy SMU among youth.

Several narrative and meta-analytic reviews have indicated small but statistically significant correlations ($r = .11$–$r = .17$) between duration of SMU and depressive symptoms (Orben, 2020). Similarly, a systematic review of 13 studies in youth found that time spent

This article was published Online First April 22, 2024.

Christopher G. Davis https://orcid.org/0000-0003-3831-080X

Gary S. Goldfield https://orcid.org/0000-0001-6216-7824

The authors report that there are no competing interests to declare. The data that support the findings of this study are available in the Open Science Framework at https://osf.io/5c847/.

Christopher G. Davis served as lead for data curation, formal analysis, project administration, and visualization. Gary S. Goldfield served as lead for supervision and validation. Christopher G. Davis and Gary S. Goldfield contributed equally to conceptualization, writing–original draft, writing–review and editing, and methodology.

Correspondence concerning this article should be addressed to Gary S. Goldfield, Healthy Active Living and Obesity Research Group, Children's Hospital of Eastern Ontario Research Institute, 401 Smyth Road, Ottawa, ON K1H 8L1, Canada. Email: ggoldfield@cheo.on.ca

on social media, repeated checking for messages, personal investment on social media, and addictive or problematic SMU was associated with more severe symptoms of depression, anxiety, and psychological distress (Keles et al., 2020).

The relatively weak associations between heavy SMU and psychological distress in youth may indicate that heavy use carries greater risk for some people than others. That is, whereas heavy SMU may not be harmful to well-adjusted, socially integrated youth, similar use may exacerbate distress in vulnerable youth (e.g., youth who feel insecure, have concerns about their body image, are anxious, or are experiencing symptoms of depression or dysphoria). Several mechanisms have been proposed. *First*, relative to socially integrated youth, vulnerable youth may be more prone to be affected by and targeted with harmful social media content (e.g., teasing, cyberbullying; Sampasa-Kanyinga & Hamilton, 2015; Sampasa-Kanyinga et al., 2018; Sampasa-Kanyinga, Lalande, & Colman, 2020). Second, being exposed to images of celebrities and peers on social media who appear more attractive and seem to live more exciting and interesting lives may lead vulnerable youth to see their own appearance and life as worse by social comparison processes (Pera, 2018; Steers et al., 2014). Additionally, there is emerging evidence to support a third mechanism rooted in displacement theory. This theory was initially developed to explain why television viewing was associated with poorer physical and emotional development (Neuman, 1988), but has been empirically applied to the social media context. Displacement theory posits that engaging in high amounts of time on social media leads to poorer mental health because it displaces time spent in mental health-promoting behaviors (Kraut et al., 1998). Indeed there is empirical support for this theory given epidemiological studies in youth show that heavy (2 hr/day or more) SMU is associated with later bedtime and reduced sleep duration and sleep quality (Sampasa-Kanyinga et al., 2018; Sampasa-Kanyinga, Lalande, & Colman, 2020), which in turn is associated with more severe symptoms of anxiety and depression (Chaput et al., 2016). Also consistent with displacement theory, there is evidence that SMU (and digital media in general) is associated with lower physical activity levels (Sampasa-Kanyinga & Chaput, 2016) and less in-person social interaction, and connectedness with peers and parents (Sampasa-Kanyinga, Goldfield, et al., 2020), factors that are known to protect against the development of psychopathology in youth. It is important to note that these mechanisms are not mutually exclusive.

To date, the evidence for the harmful effects of heavy SMU has not been compelling. As noted by Orben (2020) and Keles et al. (2020), the substantial heterogeneity in methods used across studies combined with the low-quality evidence emanating from predominantly cross-sectional designs limits conclusions drawn from this body of literature, highlighting the need for experimental studies to better understand the psychological effects of SMU.

Few experimental studies on this issue exist. Tromholt (2016) showed that eliminating Facebook use for a period of 1 week significantly improved well-being in Danish adults relative to controls. Turel et al. (2018) likewise demonstrated that short (1 week) periods of abstinence from social media reduced perceived stress, particularly among heavy users. Yet two other studies reported null effects on the well-being of abstaining from SMU for up to 4 weeks (Agadullina et al., 2020; Hall et al., 2021). Thai et al. (2021) found that limiting SMU to 1 hr/day for 3 weeks led to reductions in symptoms of anxiety but not depression in a relatively small sample of youth. In contrast, Hunt et al. (2018) found that reducing SMU on three platforms (Facebook, Snapchat, Instagram) to 10 min/day each for 3 weeks led to a reduction in depressive symptoms among youth. However, they found no intervention effects on anxiety or fear of missing out (FoMO)—the sense of apprehension that one is missing out on pleasant or enjoyable experiences that others are enjoying, a phenomenon shown to motivate greater SMU, with associations with higher distress (Przybylski et al., 2013). Although encouraging, most of these experimental studies have issues that limit their utility. For instance, Agadullina et al. (2020), Hall et al. (2019), and Turel et al. (2018) relied on self-reported compliance to SMU restriction. Self-reported SMU has been shown to be a poor predictor of actual use (Parry et al., 2021). Tromholt (2016), Agadullina et al. (2020), Hall et al. (2019), and Turel et al. (2018) required complete SMU abstinence—a goal that is unrealistic for heavy social media using youth. Hunt et al.'s (2018) restriction to 10 min/day per social media platform may seem overly rigid to youth today, who have more social media options available to them and likely value the freedom to choose how they allocate their SMU when restricted.

The inconsistent findings in the experimental studies and the relatively weak correlations found in cross-sectional studies suggest that not all heavy social media users experience ill effects. Hunt et al. (2018), for instance, found that reducing SMU had greater ameliorative effects on those who were initially depressed relative to those who were not. Of course, this could be attributable to the fact that those with more symptoms have more room to change, but it may also suggest that individuals with elevated distress who are more psychologically vulnerable may be more inclined to make unfavorable social comparisons on social media (Bäzner et al., 2006), to be passive in their use (Verduyn et al., 2017), and/or displace in-person social connections with peers and other mental health-promoting activity in favor of using social media, consistent with displacement theory (Blackwell et al., 2017).

Accordingly, the present study is designed to address the above-noted limitations by experimentally investigating whether reducing objectively measured SMU on smartphones to 1 hr/day for 3 weeks leads to a reduction in depression, anxiety, and FoMO in a large sample of youth with emotional distress. Following Thai et al. (2021), we focus on youth with emotional distress because this is a population that is at greater risk of experiencing the negative effects of heavy use of social media, thus may benefit more from its reduction. In addition to assessing the effect of reducing SMU on indicators of mental health, and guided by displacement theory of how excessive screen time adversely impacts mental health, we also consider the effect that the intervention will have on sleep, given high SMU use is related to later bedtime and reduced sleep duration in youth (Sampasa-Kanyinga et al., 2018; Sampasa-Kanyinga, Lalande, & Colman, 2020), and shorter sleep duration is associated with greater symptoms of anxiety and depression (Chaput et al., 2016).

- Primary hypotheses: Those randomly assigned to voluntarily reduce SMU to 1-hr/day will show a significant reduction in symptoms of depression (Hypothesis 1a [H1a]), anxiety (Hypothesis 1b [H1b]), and FoMO (Hypothesis 1c [H1c]), as well as increases in sleep duration (Hypothesis 1d [H1d]) relative to status quo controls.

Finally, given that women are more likely than men to experience anxiety and depression (Georgiades et al., 2019), combined with some reports of heavier SMU and more adverse associations with mental health among women (Baxter et al., 2014; Keles et al., 2020; Substance Abuse and Mental Health Services Administration, 2017), we consider the extent to which gender (men vs. women) moderates the effect of SMU reduction on mental health outcomes. Many of the correlational studies reviewed above have examined sex differences in the relation between SMU and anxiety/depression, but none of the experimental studies have tested or found gender differences, possibly owing to sample size issues.

- Secondary analyses: Does the effect of reducing social media on depression, anxiety, FoMO, and sleep differ as a function of gender (men vs. women)?

Method

Participants

Undergraduate students enrolled in introductory psychology classes at a Canadian university were recruited to participate in a study entitled "Limiting Social Media Screen-Time on iPhones and Androids." Eligibility requirements were that (a) they were regular social media users (defined as at least 2 hr/day on average), (b) they possessed and regularly used an iPhone (running on iOS 12 or later) or Android smartphone (running on Pie 9 or later), (c) they were between the ages of 17–25 years, and (d) they were experiencing at least two of four symptoms of depression and anxiety as presented on the recruitment notice. All those who were interested were told that each eligible person had a 50% chance of being assigned to the experimental condition where they would be asked to reduce their social media screen time. They were also advised that they would be asked to submit screenshots each morning showing their social media usage for the day before. Although potential participants were aware that the study was about limiting social media screen time (an ethical obligation), they were not aware of our hypotheses until they were debriefed at the study's conclusion.

Participants were recruited over three academic terms (Winter, Summer and Fall of 2021). Two hundred and seventy-nine eligible participants initially enrolled in the study. The Consolidated Standards of Reporting Trials (CONSORT) diagram in Figure 1 summarizes information on dropouts and exclusions. Our analyses are based on a sample of 220 participants (168 women, 50 men, and two indicating "other"). Participants were compensated with grade-raising credit for their introductory psychology class. This study received ethics approval by the authors' institutional review boards and all participants provided informed consent prior to enrollment.

Design

This study used a parallel-group, randomized controlled trial design and was conducted over three semesters spanning January 2021 to December 2021. This study was designed in compliance with CONSORT guidelines for nonpharmacological trials (Boutron et al., 2017). The intervention period lasted for 3 weeks following a 1-week baseline period. Participants had an equal chance of being assigned to either the intervention or the control group using a computer-generated randomization scheme. Participants in the intervention group were instructed to reduce SMU to a maximum of 1 hr/day. This SMU reduction target was based on consultations with our panel of youth with lived experience and guided by the specific, measurable, achievable, relevant, and time-bound (SMART) goal behavioral principles which state that behavior change goals are more likely to be achieved if they are specific, measurable, achievable, relevant, and time-bound. The 1 hr/day SMU goal was also informed by epidemiological data suggesting that using over 1 hr of social media per day is associated with greater emotional distress (Twenge & Campbell, 2019). The social media platforms that were targeted in this study include Facebook, Instagram, TikTok, Snapchat, Twitter, Pinterest, and Tumblr. Messaging apps (e.g., Facebook Messenger, WhatsApp, Reddit, and texts) were not targeted for reduction as research has shown these to be separate domains of social media (Figueira Jacinto & Arndt, 2018; Kuss & Griffiths, 2017). Participants in the control group monitored their daily SMU and sent daily screenshot but were instructed to use social media as usual throughout the study. Participants in both groups received a daily reminder email for the duration of the study to send their SMU screenshots for the previous day. Furthermore, participants in the intervention group who exceeded the 1 hr/day SMU limit received a reminder email about respecting the SMU limit. Participants were also positively reinforced via email for achieving their SMART goals every time the goal was achieved. These behavioral principles of SMART goal setting, cuing, and immediate positive reinforcement have been empirically shown to be effective components of lifestyle behavior change (Goldfield et al., 2002).

Procedure

As the experiment was conducted during the COVID-19 pandemic, all procedures were conducted virtually. Participants were informed orally (via Zoom) and in writing about the study purpose, its requirements, and potential risks as part of the informed consent process. After confirming eligibility and providing consent, participants were shown how to locate their smartphone's built-in social media tracking summary on their smartphone and were asked to email a test screenshot of it to the secure study inbox. They provided permission for research staff to send daily email reminders to their preferred email address reminding them to submit their daily social media usage screenshot each night. They were then directed to the online baseline questionnaire using a secure data management software, Qualtrics.

For the first 7 days (baseline period), participants were instructed to use social media as usual. On the seventh 7th day, participants were randomly assigned to either the intervention or the control condition. Beginning on the 8th day and lasting until Day 28 (3-week intervention phase), those in the intervention condition were instructed to limit their SMU to 60 min/day, while those assigned to the control group were instructed each day to use social media as usual. On the 28th day, participants were directed to complete postintervention outcome measures online using Qualtrics.

Study materials and data are available for viewing at https://osf.io/5c847/. The study was not preregistered.

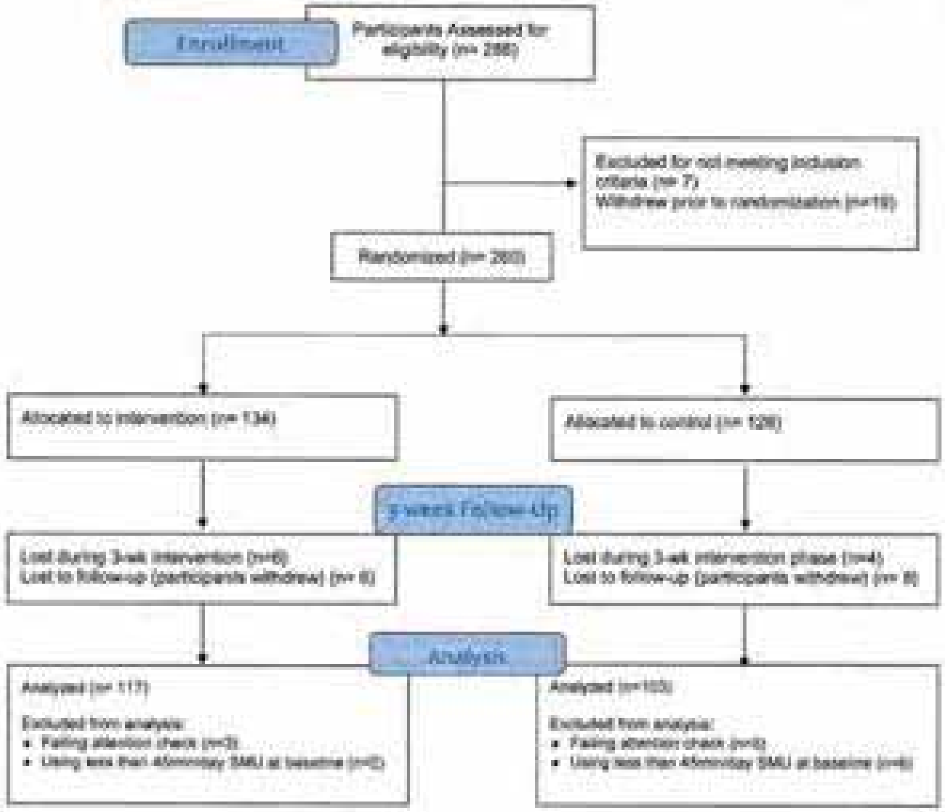
Measures

SMU

Daily SMU via smartphone was objectively assessed using integrated smartphone screen time reports and sent to the study's secured inbox every day for the duration of the study. These applications for

Figure 1

CONSORT Flow Chart of Participants Enrollment, Randomization, Allocations, and Analysis Throughout the Study Timeline



Note. wk = week; SMU = social media use; CONSORT = Consolidated Standards of Reporting Trials. See the online article for the color version of this figure.

iPhones and Android phones enable tracking of time spent on each targeted application, and this objective measurement provides greater reliability compared to self-reported measures of SMU, which are subject to recall bias (Lee et al., 2017; Parry et al., 2021).

Depressive Symptoms

Depressive symptomatology was assessed at the beginning (Day 1) and the end of the experiment (Day 28) with the revised 10-item Center for Epidemiological Studies Depression (CES-D) scale (Andresen et al., 1994; original CES-D by Radloff, 1977). This revised and shortened CES-D has been shown to be a valid and reliable measure of depressive symptoms in a variety of samples (Andresen et al., 1994; González et al., 2017; Irwin et al., 1999) including university students (Bradley et al., 2010). Scores of 10 or more are considered indicative of clinically concerning depression symptoms. Recall that we aimed to target recruitment of youth experiencing emotional distress based on the belief this population would be more vulnerable to the harms of SMU, and may therefore benefit more from its reduction than youth without distress. Although inclusion criteria required all participants to report being distressed, 70% of the sample scored above the threshold of 10 at baseline. Cronbach's α over the two time points averaged .85.

Generalized Anxiety

Generalized anxiety was measured at the beginning and end of the experiment with the GAD-7 (Spitzer et al., 2006). The seven items that make up the GAD-7 are based on *Diagnostic and Statistical Manual of Mental Disorders* (fourth edition, *DSM-IV*) criteria for generalized anxiety. The instrument has been widely used in clinical, general population, and youth and university student samples (e.g., Byrd-Bredbenner et al., 2021; Löwe et al., 2008; Tiirikainen et al., 2019) and has been validated against clinical diagnoses and comparable instruments by Spitzer et al. (2006). A systematic review and meta-analysis of the instrument's diagnostic accuracy against a structured clinical interview indicated that a cut point of 8 had pooled sensitivity and specificity >0.80 (Plummer et al., 2016). In the current study, 58% of participants scored above the clinical threshold of 8 at Time 1. Cronbach's α averaged .89.

FoMO

We used Przybylski et al., (2013) 10-item questionnaire to assess FoMO at the beginning and end of the experiment. This instrument has demonstrated its concurrent validity with positive correlations of scores on the FoMO scale with the level of social media engagement and negative correlations with the level of needs satisfaction, life

satisfaction, and (positive) mood (Przybylski et al., 2013). Moreover, scores on the FoMO scale have also been shown to correlate positively with depression and Internet-communication disorder (Wegmann et al., 2017) and with social anxiety and problematic Facebook use (Dempsey et al., 2019). Items are rated on a 5-point scale where 1 = *not at all true of me* to 5 = *extremely true of me*. We report participants' mean of the 10 items. In the present study, Cronbach's α averaged .86.

Sleep

To assess hours of sleep per night, participants were asked in both the preliminary survey and the follow-up survey the following questions separately for weekdays and weekends: "During the past week, what time have you usually turned out the lights to go to sleep on [weekdays/weekends]?" and "During the past week, what time have you usually woken up in the morning on [weekdays/weekends]?" In the preliminary survey, nine participants did not provide data on their sleep, and in the follow-up survey, 38 did not provide sleep data. Outliers (<2 hr and >14 hr) were replaced so that the minimum hours per night was 2 and maximum was 14. Weekday hours of sleep and weekend hours of sleep were comparable (averaging approximately 8 hr/night; SD = 1.55–1.64) and positively correlated in the baseline survey (r = .51) and in the follow-up survey (r = .68). As such, they were averaged at each time point.

Analysis Strategy

Manipulation Check

A 2 (condition) × 4 (week) mixed analysis of variance (ANOVA) was conducted to compare daily SMU as measured by screenshots of use among participants in each condition during the 4-week study period to evaluate the success of the intervention in limiting SMU.

Primary and Secondary Analysis

To test whether the intervention has effects on primary (depression and anxiety symptoms) and secondary outcomes (FoMO and sleep), 2 × 2 mixed ANOVAs were conducted separately for each outcome. Significant effects were explored using simple effects. We also considered whether effects differed as a function of gender by adding gender (men/women) as a third factor. Effect sizes for each ANOVA are reported in partial eta-squared, with values ranging from 0.01 to 0.05, 0.06 to 0.13, and 0.14 or higher representing small, medium, and large effects, respectively.

Power Analysis

Statistical power was calculated using general linear mixed model power and sample size 3.0.0 (Kreidler et al., 2013) software designed to assess power in longitudinal designs. Power was estimated based on a hypothesized Condition × Time interaction using a repeated measures ANOVA such that those receiving the intervention were expected to decline one third of a standard deviation in our main outcomes (depression and anxiety) whereas controls were expected to maintain their prior level, and anticipating a pre-to-post correlation in dependent variables of r = .6 (stronger test–retest correlations yield greater power). Effect sizes were estimated based on findings of Hunt et al. (2018) and Thai et al. (2021). Assuming a criterion for significance of p = .05, a sample of 200 participants was required to achieve this between group effect size with a power of 0.80.

Results

Participants in both the experimental and control groups reliably provided daily screenshots of the SMU. During the baseline period, 94.5% provided screenshots on all 7 days. In the 21 days following assignment to intervention or control condition, 93.2% provided screenshots on at least 20 days, with rates not differing by condition for either baseline (p = .71) or during intervention (p = .99). Mean daily social media screen time was calculated for each week, with participant's daily mean for that week substituted for any missing data.

Experimental and control groups did not differ significantly at baseline on any of the study variables (see Table 1).

Manipulation Check

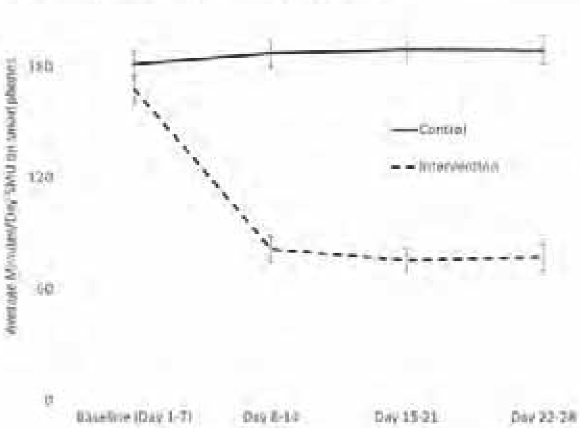
A 2 (condition) × 4 (week) ANOVA on average daily social media screen time revealed the expected condition by time (in

Table 1
Baseline Characteristics of Study Population on Demographic Values and Mental Health Indicators

Variable	Condition			Differences between groups (p)
	Grouped (n = 220)	Intervention (n = 117)	Control (n = 103)	
Gender	50M/168W/2O	25M/92W	25M/76W/2O	.55
Age				.39
17–19 years old (n)	161	84	77	
20–22 years old (n)	37	22	15	
23–25 years old (n)	11	4	7	
Baseline variable	M (SD)	M (SD)	M (SD)	
Depression	14.19 (5.66)	13.96 (5.51)	14.45 (5.84)	.52
Anxiety	9.86 (6.16)	9.43 (6.42)	10.34 (5.86)	.28
FoMO	2.68 (.84)	2.65 (.86)	2.72 (.82)	.59
Sleep (hours/day)	8.35 (1.39)	8.27 (1.49)	8.43 (1.27)	.39

Note. Missing age data of 11 participants due to technical issues. Missing sleep data for nine participants due to nonresponse. M = men; W = women; O = other; FoMO = fear of missing out.

Figure 2
Total Daily SMU Over Time by Condition



Note. Error bars represent standard errors. SMU = social media use.

weeks) interaction, $F(3, 648) = 94.048$, $p < .001$, $\eta_p^2 = .255$. Simple effects indicated no difference between intervention and control groups during the baseline period (days 1–7; $p = .197$), but significant differences by condition in each of the subsequent weeks (all $p < .001$), with those in the intervention condition averaging 78.25 min/day (reducing their daily SMU by approximately 50%) whereas those in the control condition averaging 188.76 min/day (see Figure 2).

Main Analyses

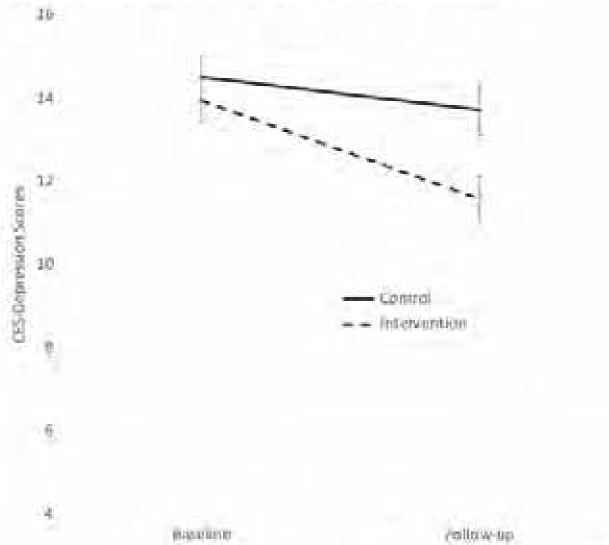
Effects of Reducing Social Media on Depressive Symptoms (H1a)

The ANOVA on depression indicated a marginal main effect of condition, $F(1, 217) = 3.63$, $p = .058$, $\eta_p^2 = .016$; a significant main effect of time, $F(1, 217) = 21.42$, $p < .001$, $\eta_p^2 = .090$; and a significant condition by time interaction, $F(1, 217) = 5.35$, $p = .022$, $\eta_p^2 = .024$. Simple effects indicated that whereas the control group did not decline significantly (decrease of 0.79, $p = .115$), the intervention group declined significantly in levels of depression (decrease of 2.36 points, $p < .001$; see Figure 3). Adding gender as a factor to the model yielded no significant effects involving gender (men vs. women) and did not significantly alter any of the effects described above.

Effects of Reducing Social Media on Anxiety Symptoms (H1b)

The ANOVA on anxiety indicated a significant main effect of condition, $F(1, 216) = 4.33$, $p = .039$, $\eta_p^2 = .020$; of time, $F(1, 216) = 37.17$, $p < .001$, $\eta_p^2 = .147$; and an interaction of condition by time, $F(1, 216) = 5.99$, $p = .015$, $\eta_p^2 = .027$. Simple effects indicated that whereas both groups declined over time, the decrease was greater for those assigned to the intervention (decrease of 2.35 points, $p < .001$) relative to controls (decrease of 1.01 points, $p = .013$; see Figure 4). Adding gender as a factor to the model yielded no significant effects involving gender and did not significantly alter any of the effects described above.

Figure 3
Effect of Reducing SMU on Symptoms of Depression

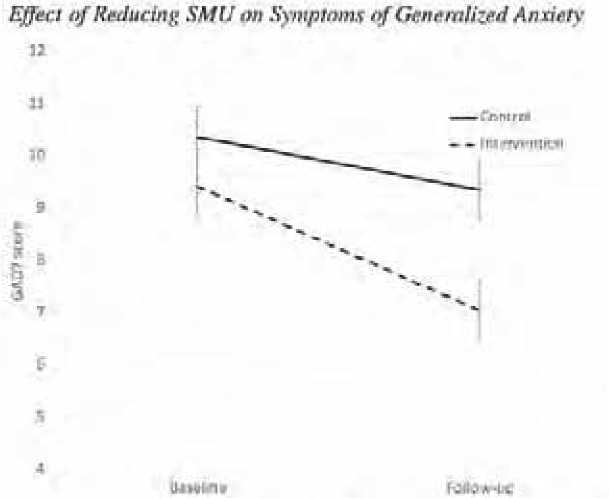


Note. Error bars represent standard errors. SMU = social media use; CES = Center for Epidemiological Studies.

Effects of Reducing Social Media on FoMO (H1c)

The ANOVA on FoMO indicated no significant main effect of condition, $F(1, 217) = 2.34$, $p = .128$, $\eta_p^2 = .011$, a significant effect of time, $F(1, 217) = 44.60$, $p < .001$, $\eta_p^2 = .170$, and a significant interaction of condition by time, $F(1, 217) = 3.95$, $p = .048$, $\eta_p^2 = .018$. Simple effects indicated that whereas both groups declined over time, the decrease was greater for those assigned to the intervention ($p < .001$) relative to controls ($p = .002$; see Figure 5). Adding gender as a factor to the model yielded no

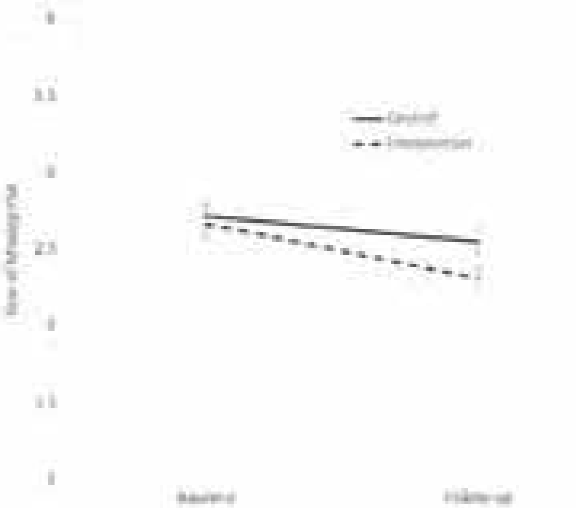
Figure 4
Effect of Reducing SMU on Symptoms of Generalized Anxiety



Note. Error bars represent standard errors. SMU = social media use; GAD-7 = generalized anxiety disorder 7.

Figure 5

Effect of Reducing SMU on Fear of Missing Out



Note. Error bars represent standard errors. SMU = social media use.

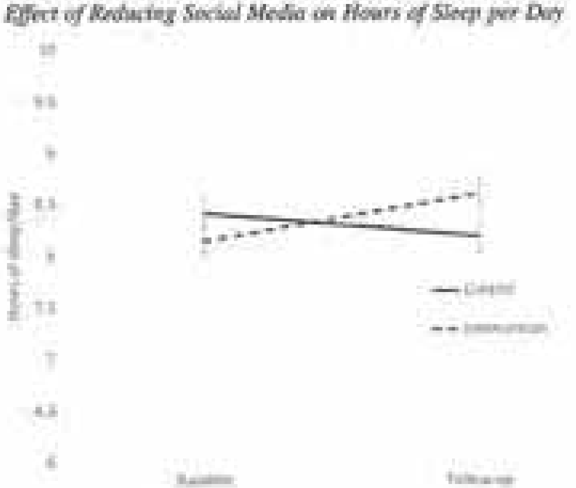
significant effects involving gender and did not significantly alter any of the effects described above.

Effects of Reducing SMU on Sleep (H1d)

The ANOVA on sleep yielded a nonsignificant main effect of time, $F(1, 176) = 1.15$, $p = .284$, $\eta_p^2 = .007$; a nonsignificant main effect of condition, $F(1, 176) = 0.14$, $p = .709$, $\eta_p^2 = .001$; and a significant interaction, $F(1, 176) = 9.52$, $p = .002$, $\eta_p^2 = .051$. Simple effects analyses indicated that whereas hours of sleep declined for control participants by approximately 15 min per night ($p = .179$), sleep increased for those in the intervention condition by about 30 min per night ($p = .002$; see Figure 6). Adding gender as a factor to the model yielded no significant effects involving gender and did not significantly alter any of the effects described above.

Figure 6

Effect of Reducing Social Media on Hours of Sleep per Day



Note. Error bars represent standard errors.

Discussion

As hypothesized, experimentally reducing SMU in youth with emotional distress to approximately 1 hr/day for 3 weeks led to significant reductions in anxiety and depression symptoms relative to self-monitoring controls who had unrestricted access to SMU. These findings add experimental evidence showing a clear causal link between SMU and mental health, consistent with a growing body of predominantly cross-sectional evidence showing that heavy use of social media may be psychologically harmful to youth (Twenge & Campbell, 2019; Woods & Scott, 2016). Unlike previous studies, however, we limited our participant pool to those who were currently reporting symptoms of anxiety or depression on the assumption that reducing use is most likely to be salubrious among those with preexisting symptoms. Youth who are not experiencing distress may not be as affected by heavy SMU, and—as Hunt et al. (2018) observed—reducing SMU among this group may render smaller mental health benefits. Moreover, this study found experimental evidence to support the displacement hypothesis in explaining how reducing social media may have a favorable impact on mental health. Specifically, reducing SMU led to increased sleep, and greater sleep is associated with lower anxiety and depression in youth (Chaput et al., 2016).

To our knowledge, our study is the first to show that SMU reduction to approximately 1-hr/day for 3 weeks led to a decrease in both anxiety and depressive symptoms in youth with emotional distress, whereas previous experimental studies in healthy populations have shown somewhat inconsistent findings (Agadullina et al., 2020; Hall et al., 2019; Hunt et al., 2018; Tromholt, 2016). There are many potential explanations for these findings. Youth with emotional distress tend to be less engaged with peers and organized activities (Steger & Kashdan, 2009), report more social isolation (Achterbergh et al., 2020), unfavorable social comparisons (Bäzner et al., 2006) and may use SMU to stay connected with peers to compensate for a lack of offline interpersonal relationships (Blackwell et al., 2017). This may make them more vulnerable to the psychologically harmful elements of heavy SMU exposure that result from the preponderance of portrayals of online profiles and posts that over-represent positive experiences, photo-edited pictures, and displays of high number of friends/followers and "likes" as a reflection of perceived online popularity and elicit envy (Pera, 2018). Thus, reducing time spent using social media may confer psychological benefits by reducing exposure to these "toxic" elements and unfavorable psychological comparison processes, which have been noted to be more frequent and harmful in SMU contexts as compared to offline environments (Lin et al., 2016; Walther et al., 2011). Regardless of the mechanisms, our findings show widespread psychological benefits of reducing SMU in a vulnerable population of youth with emotional distress.

To the extent that people use social media to stay connected with peers (especially during a pandemic when this study was conducted), one might expect that limiting one's time on social media might leave one feeling left out and isolated. Some have argued that such FoMO keeps people tied to social media (Przybylski et al., 2013), but also that high SMU leads to greater FoMO (Oberst et al., 2017). This highlights the bidirectional nature of these relationships. The drive to use smartphones is very powerful, evidenced by studies showing social media possesses greater reinforcing properties than palatable snack foods (O'Donnell & Epstein, 2019) and evokes the release of

dopamine, with concomitant activation in brain regions that are implicated in the development of drug and alcohol addiction (Montag et al., 2017; Sarıyska et al., 2018). Similarly, it has been demonstrated that the intermittent and unlimited reinforcement from receiving SMU notifications leads to habitual SMU as a means to prevent FoMO (Griffiths, 2018). Given these reinforcement processes inherent in SMU combined with the fact that we did not instruct participants to turn off their SMU notifications, it would not be unreasonable to predict that limiting access to SMU would increase FoMO, but ironically our study showed the opposite: that reducing SMU by about 90 min/day (approximate 50% reduction from baseline) for 3 weeks led to a greater reduction in FoMO relative to controls. This is a novel finding that has important public health implications given the high prevalence of excessive SMU in youth, and the reliable relationship between FoMO and emotional distress (Przybylski et al., 2013). Interestingly, our results are consistent with studies using an abstinence model that show that whereas short-term (24-hr) abstinence from SMU may temporarily increase perceived FoMO and emotional distress (Roberts & Koliska, 2014), 7-day abstinence tends to reduce FoMO and increase social connection (Brown & Kuss, 2020). This pattern of results is also consistent with addiction research indicating that the heightened distress initially experienced upon cessation or marked reduction in use (i.e., withdrawal) decreases over time with adaptation when individuals learn to meet their psychological needs in other ways (Brown & Kuss, 2020; Turel, 2015). Reducing the exposure of SMU to a moderate amount of use (about 1-hr/day) may provide an optimal balance between allowing enough SMU to feel connected to peers and meet psychological needs while sufficiently reducing exposure to the harmful effects of SMU (e.g., unfavorable social comparisons), and/or perhaps offering time for healthier pursuits, leading to reduced FoMO and emotional distress.

According to the displacement hypothesis (Neuman, 1988), spending large amounts of time on social media (or screens in general) displaces time spent on mental health-promoting behaviors like sleep, physical activity, time in nature, and recreational activities and hobbies (Guerrero et al., 2019; Nie, 2001). Until now, the evidence for this has relied on data from cross-sectional studies showing that SMU is associated with delayed bedtime, shorter sleep duration, and poorer sleep quality (Levenson et al., 2017; Sampasa-Kanyinga et al., 2018; Sampasa-Kanyinga, Lalande, & Colman, 2020; Scott et al., 2019; Woods & Scott, 2016). Additionally, systematic reviews indicate that shorter sleep duration is associated with greater anxiety and depression in youth (Chaput et al., 2016) and adults (Ross et al., 2020). To our knowledge, our study is the first to experimentally demonstrate that reducing SMU by about 90 min/day led to significant increases—30 min/night—in sleep, whereas the control group showed a reduction of 15 min/night over the course of the study, for a relative group difference of 45 min/night. This suggests that one possible mechanism by which reducing SMU may lead to reductions in emotional distress is through increased sleep, consistent with displacement theory.

Gender Differences

We examined how gender might moderate intervention effects given studies showing that women are at greater risk of anxiety and depression, tend to spend more time in SMU, and may be at greater risk of psychological harm from heavier SMU compared to men (Keles et al., 2020). We found no gender differences in the effect of limiting social media on the psychological outcomes of interest, possibly due limited power resulting from the fact that only 23% of the sample was comprised of men. This gender breakdown was somewhat expected given there are more women than men in undergraduate psychology, and women are more likely to experience emotional distress, and are heavier social media users than are men. That we found no gender differences suggests that whereas young women may be at greater risk, young men and women derive comparable psychological benefits from SMU reduction.

Strengths and Limitations

This study has many methodological strengths and weaknesses that warrant mention. Study strengths include the randomized controlled trial design with an active control group that controlled for self-monitoring of SMU, an objective measure of SMU on smartphones, good compliance to the intervention, and validated measures of mental health, all of which strengthen the internal validity of the findings. In addition, we implemented the intervention virtually at very little cost, enhancing the public health implications of the findings. Moreover, we targeted a population of youth presenting with emotional distress during a critical developmental period that puts them at risk for lifelong mental illness, making results more clinically impactful.

These strengths are balanced by several weaknesses. The intervention period only lasted 3 weeks and thus represents a proof of principle type study. Future studies should investigate whether reducing SMU over longer periods of time is a feasible and effective way for producing sustained mental health benefits in youth. Although most participants were quite compliant with the SMU restriction on their smartphones, we could not objectively clamp SMU, thus some participants in the intervention group greatly exceeded the 1-hr daily goal. Nevertheless, the intervention group still reduced SMU by approximately 50%, on average, from baseline. It should be recognized that the 1 hr/day goal was derived by cross-sectional research showing more than 1 hr of digital media use is associated with greater distress in a dose–response manner (Twenge & Campbell, 2019), this target is somewhat arbitrary, but it served as a specific, viable goal that participants could measure themselves against as opposed to something vague like "reduce to 50% of your typical SMU." Although participants did not always meet the target, they did cut their SMU substantially and consequently were likely more mindful of how they used their time on social media. Finally, our sample comprised university students who were willing to participate in a study where they knew there was a 50% chance that they would be asked to limit their SMU; they had some interest in reducing their SMU. An unselected sample may be less inclined to comply with requests to limit their use, and thus may not realize the mental health benefits.

Future Research Directions

Future research should assess the mechanisms through which reductions in SMU improve mental health. Several putative mechanisms have been proposed, including reductions in unfavorable social comparisons, reduced exposure to harmful content, and more frequent and enhanced in-person social interactions, with evidence from the current study suggestive of increased sleep. As such, results are consistent with displacement theory, but a more rigorous

test of this theory is warranted by incorporating many other health-promoting behaviors beyond sleep that could be displaced by high SMU, and by extension, substituted for when SMU is constrained.

Whereas our study shows that cutting down on SMU reduces FoMO, anxiety, and depressive symptoms, and increases sleep time, we also observed reductions in FoMO and anxiety in the control group, albeit to a lesser extent. Although we do not have evidence to account for this, it may be that these reductions are due to the control group monitoring their daily use (i.e., being made aware daily through submitting screenshots of their SMU how much time they are spending on such platforms). Another possibility has been suggested by Shrout et al. (2018), who have documented the tendency for participants in longitudinal studies to slightly elevate their initial reports, particularly on measures assessing affect. They make the case that it is more likely an initial elevation rather than attenuation at follow-up. Regardless, these trends in the control group highlight the importance of using randomized controlled designs in intervention research.

Finally, our study only targeted reductions in the duration of SMU, but research shows that how people use social media (active vs. passive, open chats vs. closed, number and type of platforms used, etc.) shows differential associations with mental health (Thorisdottir et al., 2019; Tromholt, 2016; Verduyn et al., 2017), so both quantity and quality of SMU should be taken into account when designing intervention studies.

Conclusion

To our knowledge, our results are the first to show that among youth with emotional distress, reducing SMU by about 50% from baseline produced significant reductions in symptoms of depression, anxiety, and FoMO. In addition, we also found that SMU reduction led to a significant increase in sleep, consistent with displacement theory. These beneficial effects of SMU were not moderated by gender, suggesting that men and women with emotional distress derive comparable psychological benefits from reducing SMU. These findings suggest that reducing SMU may represent a feasible, affordable, and effective strategy that should be considered for inclusion in the comprehensive management of anxiety and depression in youth with emotional distress, a high-risk population for chronic mental illness. Future research using well-controlled experimental designs is needed to empirically determine whether displacement theory provides a more comprehensive mechanistic understanding relative to other competing theories (e.g., social comparison theory) on how reducing SMU confers mental health benefits.

References

Achterbergh, L., Pitman, A., Birken, M., Pearce, E., Sno, H., & Johnson, S. (2020). The experience of loneliness among young people with depression: A qualitative meta-synthesis of the literature. *BMC Psychiatry, 20*(1), Article 415. https://doi.org/10.1186/s12888-020-02818-3

Agadullina, E. R., Lovakov, A., & Kiselnikova, N. V. (2020). Does quitting social networks change feelings of loneliness among freshmen? An experimental study. *Journal of Applied Research in Higher Education, 13*(1), 149–163. https://doi.org/10.1108/JARHE-11-2019-0283

Anderson, M., & Jiang, J. (2018, May). *Monica Anderson and Jingjing Jiang*. *Pew Research*. https://assets.pewresearch.org/wpcontent/uploads/sites/14/2018/05/31102617/PI_2018.05.31_TeensTech_FINAL.pdf

Andresen, E. M., Malmgren, J. A., Carter, W. B., & Patrick, D. L. (1994). Screening for depression in well older adults: Evaluation of a short form of the CES-D (Center for Epidemiologic Studies Depression Scale). *American Journal of Preventive Medicine, 10*(2), 77–84. https://doi.org/10.1016/S0749-3797(18)30622-6

Baxter, A. J., Scott, K. M., Ferrari, A. J., Norman, R. E., Vos, T., & Whiteford, H. A. (2014). Challenging the myth of an "epidemic" of common mental disorders: Trends in the global prevalence of anxiety and depression between 1990 and 2010. *Depression and Anxiety, 31*(6), 506–516. https://doi.org/10.1002/da.22230

Bäzner, E., Brömer, P., Hammelstein, P., & Meyer, T. D. (2006). Current and former depression and their relationship to the effects of social comparison processes. Results of an internet based study. *Journal of Affective Disorders, 93*(1–3), 97–103. https://doi.org/10.1016/j.jad.2006.02.017

Blackwell, D., Leaman, C., Tramposch, R., Osborne, C., & Liss, M. (2017). Extraversion, neuroticism, attachment style and fear of missing out as predictors of social media use and addiction. *Personality and Individual Differences, 116*, 69–72. https://doi.org/10.1016/j.paid.2017.04.039

Boutron, I., Altman, D. G., Moher, D., Schulz, K. F., Ravaud, P., & CONSORT NPT Group. (2017). CONSORT statement for randomized trials of nonpharmacologic treatments: A 2017 update and a CONSORT extension for nonpharmacologic trial abstracts. *Annals of Internal Medicine, 167*(1), 40–47. https://doi.org/10.7326/M17-0046

Bradley, K. L., Bagnell, A. L., & Brannen, C. L. (2010). Factorial validity of the Center for Epidemiological Studies Depression 10 in adolescents. *Issues in Mental Health Nursing, 31*(6), 408–412. https://doi.org/10.3109/01612840903484105

Brown, L., & Kuss, D. J. (2020). Fear of missing out, mental wellbeing, and social connectedness: A seven-day social media abstinence trial. *International Journal of Environmental Research and Public Health, 17*(12), Article 4566. https://doi.org/10.3390/ijerph17124566

Bryd-Bredbenner, C., Eck, K., & Quick, V. (2021). GAD-7, GAD-2, and GAD-mini: Psychometric properties and norms of university students in the United States. *General Hospital Psychiatry, 69*, 61–66. https://doi.org/10.1016/j.genhosppsych.2021.01.002

Chaput, J.-P., Gray, C. E., Poitras, V. J., Carson, V., Gruber, R., Olds, T., Weiss, S. K., Connor Gorber, S., Kho, M. E., Sampson, M., Belanger, K., Eryuzlu, S., Callender, L., & Tremblay, M. S. (2016). Systematic review of the relationships between sleep duration and health indicators in school-aged children and youth. *Applied Physiology, Nutrition, and Metabolism, 41*(6 Suppl. 3), S266–S282. https://doi.org/10.1139/apnm-2015-0627

Comeau, J., Georgiades, K., Duncan, L., Wang, L., & Boyle, M. H. (2019). Changes in the prevalence of child and youth mental disorders and perceived need for professional help between 1983 and 2014: Evidence from the Ontario Child Health Study. *The Canadian Journal of Psychiatry, 64*(4), 256–264. https://doi.org/10.1177/0706743719830035

Dempsey, A. E., O'Brien, K. D., Tiamiyu, M. F., & Elhai, J. D. (2019). Fear of missing out (FoMO) and rumination mediate relations between social anxiety and problematic Facebook use. *Addictive Behaviors Reports, 9*, Article 100150. https://doi.org/10.1016/j.abrep.2018.100150

Figueroa Jacinto, R., & Arndt, S. (2018). Prevalence of the use of mobile instant messaging applications: The need to assess their usage in human factors investigations. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 62*(1), 406–410. https://doi.org/10.1177/1541931218621093

Georgiades, K., Duncan, L., Wang, L., Comeau, J., & Boyle, M. H. (2019). Six-month prevalence of mental disorders and service contacts among children and youth in Ontario: Evidence from the 2014 Ontario Child Health Study. *Canadian Journal of Psychiatry, 64*(4), 246–255. https://doi.org/10.1177/0706743719830024

Goldfield, G. S., Raynor, H. A., & Epstein, L. H. (2002). Treatment of pediatric obesity. In T. A. Wadden & A. J. Stunkard (Eds.), *Handbook of obesity treatment*. (pp. 532–555). Guilford Press.

González, F., Nuñez, A., Merz, E., Brintz, C., Weitzman, O., Navas, E. L., Camacho, A., Buelna, C., Penedo, F. J., Wassertheil-Smoller, S., Perreira, K., Isasi, C. R., Choca, J., Talavera, G. A., & Gallo, L. C. (2017). Measurement properties of the Center for Epidemiologic Studies Depression Scale (CES-D 10): Findings from HCHS/SOL. *Psychological Assessment, 29*(4), 372–381. https://doi.org/10.1037/pas0000330

Griffiths, M. D. (2018). Adolescent social networking: How do social media operators facilitate habitual use? *Education and Health, 36*(3), 66–69.

Guerrero, M. D., Barnes, J. D., Chaput, J.-P., & Tremblay, M. S. (2019). Screen time and problem behaviors in children: Exploring the mediating role of sleep duration. *International Journal of Behavioral Nutrition and Physical Activity, 16*(1), Article 105. https://doi.org/10.1186/s12966-019-0862-x

Hall, J. A., Johnson, R. M., & Ross, E. M. (2019). Where does the time go? An experimental test of what social media displaces and displaced activities' associations with affective well-being and quality of day. *New Media & Society, 21*(3), 674–692. https://doi.org/10.1177/1461444818804775

Hall, J. A., Xing, C., Ross, E. M., & Johnson, R. M. (2021). Experimentally manipulating social media abstinence: Results of a four-week diary study. *Media Psychology, 24*(2), 259–275. https://doi.org/10.1080/15213269.2019.1688171

Hunt, M. G., Marx, R., Lipson, C., & Young, J. (2018). No more FOMO: Limiting social media decreases loneliness and depression. *Journal of Social and Clinical Psychology, 37*(10), 751–768. https://doi.org/10.1521/jscp.2018.37.10.751

Irwin, M., Artin, K. H., & Oxman, M. N. (1999). Screening for depression in the older adult: Criterion validity of the 10-item Center for Epidemiological Studies Depression Scale (CES-D). *Archives of Internal Medicine, 159*(15), 1701–1704. https://doi.org/10.1001/archinte.159.15.1701

Keles, B., McCrae, N., & Grealish, A. (2020). A systematic review: The influence of social media on depression, anxiety and psychological distress in adolescents. *International Journal of Adolescence and Youth, 25*(1), 79–93. https://doi.org/10.1080/02673843.2019.1590851

Kraut, R., Patterson, M., Lundmark, V., Kiesler, S., Mukophadhyay, T., & Scherlis, W. (1998). Internet paradox: A social technology that reduces social involvement and psychological well-being? *American Psychologist, 53*(9), 1017–1031. https://doi.org/10.1037/0003-066X.53.9.1017

Kreidler, S. M., Muller, K. E., Grunwald, G. K., Ringham, B. M., Coker-Dukowitz, Z. T., Sakhadeo, U. R., Barón, A. E., & Glueck, D. H. (2013). GLIMMPSE: Online power computation for linear models with and without a baseline covariate. *Journal of Statistical Software, 54*(10), Article i10. https://doi.org/10.18637/jss.v054.i10

Kuss, D., & Griffiths, M. (2017). Social networking sites and addiction: Ten lessons learned. *International Journal of Environmental Research and Public Health, 14*(3), Article 311. https://doi.org/10.3390/ijerph14030311

Lee, H., Ahn, H., Nguyen, T. G., Choi, S.-W., & Kim, D. J. (2017). Comparing the self-report and measured smartphone usage of college students: A pilot study. *Psychiatry Investigation, 14*(2), Article 198. https://doi.org/10.4306/pi.2017.14.2.198

Levenson, J. C., Shensa, A., Sidani, J. E., Colditz, J. B., & Primack, B. A. (2017). Social media use before bed and sleep disturbance among young adults in the United States: A nationally representative study. *Sleep, 40*(9), Article zsx113. https://doi.org/10.1093/sleep/zsx113

Lin, L. Y., Sidani, J. E., Shensa, A., Radovic, A., Miller, E., Colditz, J. B., Hoffman, B. L., Giles, L. M., & Primack, B. A. (2016). Association between social media use and depression among US young adults. *Depression and Anxiety, 33*(4), 323–331. https://doi.org/10.1002/da.22466

Löwe, B., Decker, O., Müller, S., Brähler, E., Schellberg, D., Herzog, W., & Herzberg, P. Y. (2008). Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. *Medical Care, 46*(3), 266–274. https://doi.org/10.1097/MLR.0b013e318160d093

Montag, C., Markowetz, A., Blaszkiewicz, K., Andone, I., Lachmann, B., Sariyska, R., Trendafilov, B., Eibes, M., Kolb, J., Reuter, M., Weber, B., & Markett, S. (2017). Facebook usage on smartphones and gray matter volume of the nucleus accumbens. *Behavioural Brain Research, 329*, 221–228. https://doi.org/10.1016/j.bbr.2017.04.035

Neuman, S. B. (1988). The displacement effect: Assessing the relation between television viewing and reading performance. *Reading Research Quarterly, 23*(4), 414–440. https://doi.org/10.2307/747641

Nie, N. H. (2001). Sociability, interpersonal relations, and the Internet. *American Behavioral Scientist, 45*(3), 420–435. https://doi.org/10.1177/00027640121957277

Oberst, U., Wegmann, E., Stodt, B., Brand, M., & Chamarro, A. (2017). Negative consequences from heavy social networking in adolescents: The mediating role of fear of missing out. *Journal of Adolescence, 55*(1), 51–60. https://doi.org/10.1016/j.adolescence.2016.12.008

O'Donnell, S., & Epstein, L. H. (2019). Smartphones are more reinforcing than food for students. *Addictive Behaviors, 90*, 124–133. https://doi.org/10.1016/j.addbeh.2018.10.018

Orben, A. (2020). Teenagers, screens and social media: A narrative review of reviews and key studies. *Social Psychiatry and Psychiatric Epidemiology, 55*(4), 407–414. https://doi.org/10.1007/s00127-019-01825-4

Parry, D. A., Davidson, B. I., Sewall, C. J. R., Fisher, J. T., Mieczkowski, H., & Quintana, D. S. (2021). A systematic review and meta-analysis of discrepancies between logged and self-reported digital media use. *Nature Human Behaviour, 5*(11), 1535–1547. https://doi.org/10.1038/s41562-021-01117-5

Paus, T., Keshavan, M., & Giedd, J. N. (2008). Why do many psychiatric disorders emerge during adolescence? Nature reviews. *Neuroscience, 9*(12), 947–957. https://doi.org/10.1038/nrn2513

Pera, A. (2018). Psychopathological processes involved in social comparison, depression, and envy on Facebook. *Frontiers in Psychology, 9*, Article 22. https://doi.org/10.3389/fpsyg.2018.00022

Plummer, F., Manea, L., Trepel, D., & McMillan, D. (2016). Screening for anxiety disorders with the GAD-7 and GAD-2: A systematic review and diagnostic meta-analysis. *General Hospital Psychiatry, 39*, 24–31. https://doi.org/10.1016/j.genhosppsych.2015.11.005

Przybylski, A. K., Murayama, K., DeHaan, C. R., & Gladwell, V. (2013). Motivational, emotional, and behavioral correlates of fear of missing out. *Computers in Human Behavior, 29*(4), 1841–1848. https://doi.org/10.1016/j.chb.2013.02.014

Radloff, L. S. (1977). The CES-D Scale. *Applied Psychological Measurement, 1*(3), 385–401. https://doi.org/10.1177/014662167700100306

Roberts, J., & Koliska, M. (2014). The effects of ambient media: What unplugging reveals about being plugged in. *First Monday, 19*(8). https://doi.org/10.5210/fm.v19i8.5220

Ross, R., Chaput, J.-P., Giangregorio, L. M., Janssen, I., Saunders, T. J., Kho, M. E., Poitras, V. J., Tomasone, J. R., El-Kotob, R., McLaughlin, E. C., Duggan, M., Carrier, J., Carson, V., Chastin, S. F., Latimer-Cheung, A. E., Chulak-Bozzer, T., Faulkner, G., Flood, S. M., Gazendam, M. K., … Tremblay, M. S. (2020). Canadian 24-hour movement guidelines for adults aged 18–64 years and adults aged 65 years or older: An integration of physical activity, sedentary behaviour, and sleep. *Applied Physiology, Nutrition, and Metabolism, 45*(10 Suppl. 2), S57–S102. https://doi.org/10.1139/apnm-2020-0467

Sampasa-Kanyinga, H., & Chaput, J.-P. (2016). Use of social networking sites and adherence to physical activity and screen time recommendations in adolescents. *Journal of Physical Activity & Health, 13*(5), 474–480. https://doi.org/10.1123/jpah.2015-0343

Sampasa-Kanyinga, H., Chaput, J.-P., & Hamilton, H. A. (2019). Social media use, school connectedness, and academic performance among adolescents. *The Journal of Primary Prevention, 40*(2), 189–211. https://doi.org/10.1007/s10935-019-00543-6

Sampasa-Kanyinga, H., Goldfield, G. S., Kingsbury, M., Clayborne, Z., & Colman, I. (2020). Social media use and parent–child relationship: A

cross-sectional study of adolescents. *Journal of Community Psychology, 48*(3), 793–803. https://doi.org/10.1002/jcop.22293

Sampasa-Kanyinga, H., & Hamilton, H. A. (2015). Social networking sites and mental health problems in adolescents: The mediating role of cyberbullying victimization. *European Psychiatry, 30*(8), 1021–1027. https://doi.org/10.1016/j.eurpsy.2015.09.011

Sampasa-Kanyinga, H., Hamilton, H. A., & Chaput, J.-P. (2018). Use of social media is associated with short sleep duration in a dose-response manner in students aged 11 to 20 years. *Acta Paediatrica, 107*(4), 694–700. https://doi.org/10.1111/apa.14210

Sampasa-Kanyinga, H., Lalande, K., & Colman, I. (2020). Cyberbullying victimisation and internalising and externalising problems among adolescents: The moderating role of parent-child relationship and child's sex. *Epidemiology and Psychiatric Sciences, 29*, Article e8. https://doi.org/10.1017/S2045796018000653

Sariyska, R., Rathner, E.-M., Baumeister, H., & Montag, C. (2018). Feasibility of linking molecular genetic markers to real-world social network size tracked on smartphones. *Frontiers in Neuroscience, 12*, Article 945. https://doi.org/10.3389/fnins.2018.00945

Scott, H., Biello, S. M., & Woods, H. C. (2019). Social media use and adolescent sleep patterns: Cross-sectional findings from the UK millennium cohort study. *BMJ Open, 9*(9), Article e031161. https://doi.org/10.1136/bmjopen-2019-031161

Shrout, P. E., Stadler, G., Lane, S. P., McClure, M. J., Jackson, G. L., Clavél, F. D., Iida, M., Gleason, M. E. J., Xu, J. H., & Bolger, N. (2018). Initial elevation bias in subjective reports. *Proceedings of the National Academy of Sciences, 115*(1), E15–E23. https://doi.org/10.1073/pnas.1712277115

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine, 166*(10), 1092–1097. https://doi.org/10.1001/archinte.166.10.1092

Statistics Canada. (2019). *Physical activity and screen time among Canadian children and youth, 2016 and 2017*.

Statistics Canada. (2021, September 27). Survey on COVID-19 and mental health February to May 2021. *The Daily*. https://www150.statcan.gc.ca/n1/en/daily-quotidien/210927/dq210927a-eng.pdf?st=mrwSNSXM

Steers, M.-L. N., Wickham, R. E., & Acitelli, L. K. (2014). Seeing everyone else's highlight reels: How Facebook usage is linked to depressive symptoms. *Journal of Social and Clinical Psychology, 33*(8), 701–731. https://doi.org/10.1521/jscp.2014.33.8.701

Steger, M. F., & Kashdan, T. B. (2009). Depression and everyday social activity, belonging, and well-being. *Journal of Counseling Psychology, 56*(2), 289–300. https://doi.org/10.1037/a0015416

Substance Abuse and Mental Health Services Administration. (2017). *Results from the 2017 National survey on drug use and health*.

Thai, H., Davis, C. G., Stewart, N., Gunnell, K. E., & Goldfield, G. S. (2021). The effects of reducing social media use on body esteem among transitional-aged youth. *Journal of Social and Clinical Psychology, 40*(6), 481–507. https://doi.org/10.1521/jscp.2021.40.6.481

Thorisdottir, I. E., Sigurvinsdottir, R., Asgeirsdottir, B. B., Allegrante, J. P., & Sigfusdottir, I. D. (2019). Active and passive social media use and symptoms of anxiety and depressed mood among Icelandic adolescents. *Cyberpsychology, Behavior, and Social Networking, 22*(8), 535–542. https://doi.org/10.1089/cyber.2019.0079

Tiirikainen, K., Haravuori, H., Ranta, K., Kaltiala-Heino, R., & Marttunen, M. (2019). Psychometric properties of the 7-item Generalized Anxiety Disorder Scale (GAD-7) in a large representative sample of Finnish adolescents. *Psychiatry Research, 272*, 30–35. https://doi.org/10.1016/j.psychres.2018.12.004

Tromholt, M. (2016). The Facebook experiment: Quitting Facebook leads to higher levels of well-being. *Cyberpsychology, Behavior and Social Networking, 19*(11), 661–666. https://doi.org/10.1089/cyber.2016.0259

Turel, O. (2015). Quitting the use of a habituated hedonic information system: A theoretical model and empirical examination of Facebook users. *European Journal of Information Systems, 24*(4), 431–446. https://doi.org/10.1057/ejis.2014.19

Turel, O., Cavagnaro, D. R., & Meshi, D. (2018). Short abstinence from online social networking sites reduces perceived stress, especially in excessive users. *Psychiatry Research, 270*, 947–953. https://doi.org/10.1016/j.psychres.2018.11.017

Twenge, J. M., & Campbell, W. K. (2019). Media use is linked to lower psychological well-being: Evidence from three datasets. *The Psychiatric Quarterly, 90*(2), 311–331. https://doi.org/10.1007/s11126-019-09630-7

Verduyn, P., Ybarra, O., Résibois, M., Jonides, J., & Kross, E. (2017). Do social network sites enhance or undermine subjective well-being? A critical review. *Social Issues and Policy Review, 11*(1), 274–302. https://doi.org/10.1111/sipr.12033

Walther, J. B., Liang, Y. J., DeAndrea, D. C., Tong, S. T., Carr, C. T., Spottswood, E. L., & Amichai-Hamburger, Y. (2011). The effect of feedback on identity shift in computer-mediated communication. *Media Psychology, 14*(1), 1–26. https://doi.org/10.1080/15213269.2010.547832

Wegmann, E., Oberst, U., Stodt, B., & Brand, M. (2017). Online-specific fear of missing out and internet-use expectancies contribute to symptoms of internet-communication disorder. *Addictive Behaviors Reports, 5*, 33–42. https://doi.org/10.1016/j.abrep.2017.04.001

Woods, H. C., & Scott, H. (2016). #Sleepyteens: Social media use in adolescence is associated with poor sleep quality, anxiety, depression and low self-esteem. *Journal of Adolescence, 51*(1), 41–49. https://doi.org/10.1016/j.adolescence.2016.05.008

Received May 10, 2023
Revision received January 10, 2024
Accepted February 26, 2024 ■

Psychology of Popular Media

ISSN: 2689-6567

https://doi.org/10.1037/ppm0000560

Correction to "Limiting Social Media Use Decreases Depression, Anxiety, and Fear of Missing Out in Youth With Emotional Distress: A Randomized Controlled Trial" by Davis and Goldfield (2024)

In the article "Limiting Social Media Use Decreases Depression, Anxiety, and Fear of Missing Out in Youth With Emotional Distress: A Randomized Controlled Trial" by Christopher G. Davis and Gary S. Goldfield (*Psychology of Popular Media*, 2025, Vol. 14, No. 1, pp. 1–11. https://doi.org/10.1037/ppm0000536), the mean and standard deviation in Table 1 for the sleep (hour/day) variable for the intervention and control conditions should instead appear as 8.27 (1.49) and 8.43 (1.27), respectively. All versions of this article have been corrected.

https://doi.org/10.1037/ppm0000560

Addictive Behaviors 106 (2020) 106364

Contents lists available at ScienceDirect

Addictive Behaviors

journal homepage: www.elsevier.com/locate/addictbeh

ELSEVIER

ADDICTIVE BEHAVIORS

Investigating the links between fear of missing out, social media addiction, and emotional symptoms in adolescence: The role of stress associated with neglect and negative reactions on social media

M.A. Fabris, D. Marengo*, C. Longobardi, M. Settanni

Department of Psychology, University of Turin, Italy

HIGHLIGHTS

- Fear of missing out (FoMO) is associated with a decrease in emotional well-being in adolescents.
- We explore the mediating role of stress related neglect and negative reactions by social media peers.
- Adolescents high in FoMO experience heightened stress associated with neglect by online peers.
- Stress related to peer neglect (SS-N) is found to predict social media addiction (SMA).
- FoMO shows an indirect effect on emotional distress via both SMA and SS-N.

ARTICLE INFO

Keywords:

FoMO

ABSTRACT

...sing out (FoMO) is known to be associated with a decrease in emotional well-being in adolescents. ...ew studies have investigated the possible mediating factors between FoMO and emotional symptoms. ...ly, we studied the relationship between FoMO and emotional symptoms in a sample of 472 Italian ... aged 11–19. In particular, the study investigated the possible mediating role of perceived stress ... with experiences of neglect and negative reactions by other social media users, and social media ... self-report measures were used. Results show that FoMO directly and indirectly predicts emotional ... Additionally, FoMO is associated with increased sensitivity to stress associated with experiences of ... negative reactions by online peers, and social media addiction. Sensitivity to stress associated with ... not to negative reactions) by online peers is found to mediate the relationship between FoMO and ... addiction, which, in turn, mediates the relationship with emotional symptoms. In general, the study ... FoMO is a factor in experiencing higher sensitivity to stress associated with neglect by online peers, ...n my act as a trigger for social media addiction, and ultimately showing a negative impact on ...ell-being of adolescents. Limits and future directions for research are discussed.

...y lives, and ...up of users ...Robb, 2018; ...7). Through ... their contacts and extend their network of knowledge or find and exchange information and materials, but also contribute to the construction of their social identity in relation to peer groups, especially in terms of popularity and therefore acceptance and sense of belonging (Barker, 2009; Badenes-Ribera, Fabris, Gastaldi, Prino, & Longobardi, 2019). However, research also show significant associations exist between adolescents' social media use and Fear of Missing Out (FoMO), a construct which can be defined as a pervasive apprehension that others might be having rewarding experiences from which one is absent, and it is characterized by the desire to stay continually connected with what others are doing (Przybylski, Murayama, DeHaan, & Gladwell, 2013). FoMO can be conceptualized from the perspective of self-determination theory (SDT; Deci & Ryan, 1985), and specifically in relation with the psychological

* Corresponding author at: Department of Psychology, University of Turin, Turin, Via Verdi 10, 10124 TO, Italy.
E-mail addresses: mattoangelo.fabris@unito.it (M.A. Fabris), davide.marengo@unito.it (D. Marengo), Claudio.longobardi@unito.it (C. Longobardi), michele.settanni@unito.it (M. Settanni).

https://doi.org/10.1016/j.addbeh.2020.106364
Received 26 November 2019; Received in revised form 11 February 2020; Accepted 22 February 2020
Available online 27 February 2020
0306-4603/

Addictive Behaviors 106 (2020) 106364

Contents lists available at ScienceDirect

Addictive Behaviors

journal homepage: www.elsevier.com/locate/addictbeh

Investigating the links between fear of missing out, social media addiction, and emotional symptoms in adolescence: The role of stress associated with neglect and negative reactions on social media

M.A. Fabris, D. Marengo*, C. Longobardi, M. Settanni

Department of Psychology, University of Turin, Italy

HIGHLIGHTS

- Fear of missing out (FoMO) is associated with a decrease in emotional well-being in adolescents.
- We explore the mediating role of stress related neglect and negative reactions by social media peers.
- Adolescents high in FoMO experience heightened stress associated with neglect by online peers.
- Stress related to peer neglect (SS-N) is found to predict social media addiction (SMA).
- FoMO shows an indirect effect on emotional distress via both SMA and SS-N.

ARTICLE INFO

Keywords:
Stress
Neglect
Emotional symptoms
Adolescence
Social media addiction

ABSTRACT

Fear of missing out (FoMO) is known to be associated with a decrease in emotional well-being in adolescents. However, few studies have investigated the possible mediating factors between FoMO and emotional symptoms. In this study, we studied the relationship between FoMO and emotional symptoms in a sample of 472 Italian adolescents aged 11–19. In particular, the study investigated the possible mediating role of perceived stress associated with experiences of neglect and negative reactions by other social media users, and social media addiction. Self-report measures were used. Results show that FoMO directly and indirectly predicts emotional symptoms. Additionally, FoMO is associated with increased sensitivity to stress associated with experiences of neglect and negative reactions by online peers, and social media addiction. Sensitivity to stress associated with neglect (but not to negative reactions) by online peers is found to mediate the relationship between FoMO and social media addiction, which, in turn, mediates the relationship with emotional symptoms. In general, the study shows that FoMO is a factor in experiencing higher sensitivity to stress associated with neglect by online peers, which in turn my act as a trigger for social media addiction, and ultimately showing a negative impact on emotional well-being of adolescents. Limits and future directions for research are discussed.

1. Introduction

The use of social media has become part of our daily lives, and adolescents and young adults seem to be the largest group of users (Marengo, Longobardi, Fabris, & Settanni, 2018; Kidemar & Robb, 2018; Settanni, Marengo, Fabris, & Longobardi, 2018; Kemp, 2017). Through social media, teenagers can maintain their contacts and extend their network of knowledge or find and exchange information and materials, but also contribute to the construction of their social identity in relation to peer groups, especially in terms of popularity and therefore acceptance and sense of belonging (Barker, 2009; Badenes-Ribera, Fabris, Gastaldi, Prino, & Longobardi, 2019). However, research also show significant associations exist between adolescents' social media use and Fear of Missing Out (FoMO), a construct which can be defined as a pervasive apprehension that others might be having rewarding experiences from which one is absent, and it is characterized by the desire to stay continually connected with what others are doing (Przybylski, Murayama, DeHaan, & Gladwell, 2013). FoMO can be conceptualized from the perspective of self-determination theory (SDT; Deci & Ryan, 1985), and specifically in relation with the psychological

* Corresponding author at: Department of Psychology, University of Turin, Turin, Via Verdi 10, 10124 TO, Italy.
E-mail addresses: mariangela.fabris@unito.it (M.A. Fabris), davide.marengo@unito.it (D. Marengo), Claudio.longobardi@unito.it (C. Longobardi), michele.settanni@unito.it (M. Settanni).

https://doi.org/10.1016/j.addbeh.2020.106364
Received 26 November 2019; Received in revised form 11 February 2020; Accepted 22 February 2020
Available online 27 February 2020

need for connectedness with others as a factor in effective self-regulation and psychological well-being (Deci & Ryan, 1985). Evidence seems to indicate that adolescents with high levels of FoMO tend to make greater use of social media in order to compensate for these psychological needs (Oberst, Wegmann, Stodt, Brand, & Chamarro, 2017). As such, it is possible that adolescents with higher levels of FoMO may be at greater risk for excessive social media use (Al-Menayes, 2016; Blackwell, Leaman, Tramposch, Osborne, & Liss, 2017; Dhir, Yossatorn, Kaur, & Chen, 2018; Franchina, Vanden Abeele, van Rooij, Lo Coco, & De Marez, 2018). Several pieces of evidence suggest that FoMO may ultimately have an impact on the well-being of individuals, increasing negative affect and emotional symptoms (Baker, Krieger, & LeRoy, 2016; Milyavskaya, Saffran, Hope, & Koestner, 2018). Adolescents high in FOMO may be exposed to an increase in emotional symptoms because of the heightened feeling that they do not belong and that they are missing out on important shared experiences (Oberst et al., 2017), as well as a consequence of social media fatigue (Dhir et al., 2018). However, the mechanisms that connect FoMO to emotional symptoms are still largely unknown and poorly investigated, or only hypothesized.

One way FoMO appears to be linked to excessive use of social media is by way of influencing metacognitions concerning the importance of social media use for maintaining social relationship (Casale, Caplan, & Fioravanti, 2016; 2018). In view of this, it is reasonable to expect that adolescents high in FoMO may be particularly sensitive and more prone to distress due to experiencing neglect and negative reactions by peers on social media (Beyens, Frison, & Eggermont, 2016). In the online social media environment, indicators such as the number of received Likes, comments, and followers (Nesi, Choukas-Bradley, & Prinstein, 2018) are usually assumed as a measure of one's popularity and degree of acceptance by his/her immediate social network. For teens, and especially those high in FoMO, the experience of being excluded or ignored online or receiving negative comments could be a particularly stressful experience and, in turn, result in a decrease in perceived emotional well-being (Beyens et al., 2016). Some evidence suggests that when teenagers receive positive or negative comments about posted content, this affects their well-being and self-esteem (Valkenburg, Peter, & Schouten, 2006). The fear of not receiving comments/Likes (i.e., online neglect), or the fear of receiving negative reactions might trigger compulsive use of social media in order to fulfill their unsatisfied need to connect with others and maintain a positive online social status, e.g., by improving social media metrics (Marengo, Poletti, & Settanni, 2020). In this way, an adolescent with a high level of FoMO, as aforementioned, may be considered more of a risk to develop symptoms of addiction to social media use. These considerations seem in line with the "Interaction of Person-Affect-Cognition-Execution" (I-PACE) model for addictive behaviors recently proposed by Brand and colleagues (2016, 2019). This model is based on psychological and neurobiological variables that are potentially involved in the development of different forms of addiction. The model foresees an interaction between predisposing variables (such as psychological and psychopathological characteristics, genetic and biopsychological factors, and social cognitions) and the subjective perception of situational factors. The stress deriving from the subjective assessment of situational factors can activate 'affective, cognitive, and behavioral responses' that can contribute to the development and maintenance of specific forms of Internet-related behavioral addictions, such as the behavioral addiction to social media use. As regards the latter, it is worthy to note that debate exists in the literature concerning the operationalization of the behavioral addiction to social media use as an independent construct, as well as concerning the terminology used to refer to the construct itself, a debate that stems in part from its overlap with other Internet-related behavioral addictions (Bányai et al., 2017). Indeed, a variety of terms are used in the literature to refer to the construct, including *Social Media Addiction, Problematic Social Media Use, Social Media Overuse, Social Media Use Disorder*, and *Social Networking Use Disorder*, (e.g., Andreassen et al., 2016; Moretag, Wegmann, Sariyska, Demetrovics, & Brand, 2019; Van Den Eijnden, Lemmens, & Valkenburg, 2016). For the purpose of the present study, we refer to the construct of *Social Media Addiction* (SMA) as operationalized by the Bergen Social Media Addiction Scale (BSMAS; Andreassen et al., 2016). The BSMAS assesses six core addiction components derived from the model of behavioral addiction proposed by Griffiths (2005), namely mood modification (i.e., social media is used to promote a positive change in emotional states), salience (i.e., behavioral, cognitive, and emotional preoccupation with social media use), tolerance (i.e., need to increase use of social media over time), withdrawal symptoms (i.e., experiencing unpleasant physical and emotional symptoms when social media usage is restricted or stopped), conflict (i.e., problems ensuing because of social media usage), and relapse (i.e., addicts quickly reverting back to an excessive social media use after a period of abstinence).

The prevalence of adolescents at risk of SMA is difficult to estimate due to a variety of methodological factors and possible cultural influences, but also because of the aforementioned lack of consensus in defining SMA. Estimates range from 2.8% to 47% according to different studies, with females tending to be at greater risk (Andreassen, Pallesen, & Griffiths, 2017; Bányai et al., 2017). Studies concerning the Italian context report similar findings, with adolescent females showing an increased risk of SMA when compared with their male peers (Monacis, De Palo, Griffiths, & Sinatra, 2017), while the opposite is typically found for addiction to online gaming (Monacis, De Palo, Griffiths, & Sinatra, 2016).

Although the concept of SMA does not attract broad consensus (Bányai et al., 2017), it is known that excessive use of social media tends to affect the social functioning of the individual (Andreassen et al., 2016), and to correlate with a decrease in psychological well-being measures in adolescents. In fact, measures of SMA in adolescents have been found to be associated with depressive symptoms (Bányai et al., 2017; Dhir et al., 2018; Kircaburun et al., 2019; Pontes, 2017; Raudsepp & Kais, 2019; Woods, McIntyre, Bentall, & Corcoran, 2018), anxiety (Dhir et al., 2018; Pontes, 2017), low self-esteem (Bányai et al., 2017), and general psychological distress (Pontes, 2017). The causal relationship between SMA and emotional symptoms in adolescents is still unclear, and there is likely to be a two-way relationship between the two constructs (Cerniglia, Griffiths, et al., 2019; Cerniglia, Guicciardi et al., 2019; Ciarrochi et al., 2016; Li et al., 2018; Raudsepp & Kais, 2019). Still, a recent longitudinal study (Li et al., 2018) indicate that teens who overuse social networks tend to develop more depressive symptoms than teens who did not overuse over time.

Considering that adolescence is a critical period both for the emergence of forms of behavioral dependence, such as SMA, and for the exacerbation of emotional symptoms (Raudsepp & Kais, 2019), it is important to understand the factors involved in order to support possible prevention and intervention strategies. In this light, the goal of our work is to study the relationship between FoMO levels in adolescents and risk of exacerbating emotional symptoms. In particular, we investigate the mediating role of sensitivity to stress associated with experience of neglect experiences and negative reactions in social media and social media addiction in the relationship between FoMO and SMA. Considering previous literature, we hypothesize that FoMO might be positively associated with emotional symptoms in adolescents, both directly and indirectly. Based on the model we want to test here, we expected that teens with high FoMO levels might show a higher sensitivity to stress associated with neglect experiences and negative reactions by online peers on social media. We also hypothesize that those individuals reporting higher sensitivity to stress associated with neglect experiences and negative reactions by online peers may be at increased risk for increases the SMA, and, indirectly, might be more prone to suffer for heightened emotional symptoms.

Note. Values inside parenthesis are 95% confidence intervals. Dashed lines indicate non-significant paths ($p \geq .05$).

Fig. 1. Diagram of the mediation model with estimated path coefficients.

2. Method

2.1. Participants

The sample consisted of 472 adolescents (50% males, mean age (SD) = 13.50 (1.87)) with ages ranging from 11 to 19 attending 5 middle schools located in Northern Italy. School principals and teachers provided authorization for the participation of each class taking part in the study. Prior to data collection, student consent for participation, as well as parental consent, was obtained. Participants were informed of the nature and objectives of the study, in compliance with the ethical code of the Italian Association for Psychology (AIP). The research was approved by the university institutional review board (n. 182567).

2.2. Instruments

2.2.1. Fear of missing out

We used an adaptation to Italian of the FoMO scale (Przybylski et al., 2013). The FoMO scale consists of 10 items in the form of statements about fears, worries, and anxiety adolescents may have in relation to being out of touch with events, experiences, and conversations involving peers in their immediate social circle. The following are some sample items: "I fear others have more rewarding experiences than me," "I get anxious when I don't know what my friends are up to," and "It bothers me when I miss an opportunity to meet up with friends." A total score can be obtained by summing the items' scores, with higher total scores indicating higher FoMO.

2.2.2. Sensitivity to stress associated with neglect and negative reactions by online peers

We administered a newly devised instrument assessing two aspects of adolescents' experience on social media, namely their self-reported sensitivity to stress associated with experiences of neglect by other users (SS-N, 4 items, " I would feel stressed if my posts did not receive comments," " I would feel stressed if my pictures and videos did not receive comments," "I would feel stressed if my posts did not receive Likes," "I would feel stressed if my pictures and videos did not receive Likes"), and negative reactions by other users (SS-NeR, 4 items, "I would feel stressed if my posts received negative comments," "I feel would feel stressed if my pictures or videos received negative comments," "I would feel stressed if I got kicked out from social media groups", "I would feel stressed if I lost friends/followers on social media"). Items were rated on a 5-point scale (1 = disagree completely, 5 = completely agree). Confirmatory factor analysis (CFA) supported a two-dimensional structure of the scale: $\chi 2$ (14) = 30.590, $p < 0.001$; CFI = 0.996; RMSEA = 0.048 (prob. of RMSEA < 0.05 = 0.53).

2.2.3. Social media addiction

We administered the Italian Bergen Social Media Addiction Scale (BSMAS, Andreassen, Torsheim, Brunborg, & Pallesen, 2012, Monacis, Palo, Griffiths, & Sinatra, 2016). The BSMAS is comprised of six items assessing the following components: salience, tolerance, mood modification, relapse, withdrawal symptoms, and conflict. The items, are rated on a 5-point scale (1 = very rarely, 5 = very often), and can be summed to obtain a total score.

2.2.4. Emotional symptoms

We used the Emotional Symptoms subscale from the Italian (Marzocchi et al., 2004) version of the Strengths and Difficulties Questionnaire (SDQ; Goodman, Meltzer, & Bailey, 1998) to assess students' self-reported perceptions of their own emotional distress. The SDQ consists of 25 items measuring 5 dimensions: Emotional Symptoms, Conduct Problems, Hyperactivity/Inattention, Peer Relationship Problems, and Prosocial Behavior. All items are rated on a three-point Likert scale (0-not true to 2-certainly true), and raw scores are used to compute the five subscale scores; a higher score indicates more difficulties or strengths, depending on the subscale. As regards the Emotional Symptoms, the subscale includes 5 items (e.g., "I am often unhappy",4 "I worry a lot," "I have many fears, I am easily scared").

2.3. Data analysis

First, we inspect the reliability of administered instruments using Cronbach's alpha reliability coefficient. Then, we compute descriptive statistics (mean, standard deviation for continuous measures, and percentage for gender), as well as Pearson's correlations between all study measures. Next, we investigate the interplay between FoMO, SS-N, SS-NeR, and SMA in predicting ES. More specifically, we perform a path analysis via a set of multiple regression analyses to estimate the path coefficients shown in Fig. 1. That is, we investigate the following direct effects: 1) the direct effect of FoMO on SS-N and SS-NeR; 2) the direct effects of FoMO, SS-N, and SS-NeR on SMA; 3) the direct effects of FoMO, SS-N, SS-NER, SMA on ES. Additionally, we investigate the following indirect effects: 1) the separate, simple indirect effects of FoMO on ES passing through SS-N, SS-NeR, and BSMAS; 2) the serial indirect effect of FoMO on ES passing through both SS-N and SMA; 3) the serial indirect effect of FoMO on ES passing through both SS-NeR and SMA. Direct and indirect effects and relative 95% confidence intervals are estimated using 5000 bootstrap samples. Note that in estimating all path coefficients, we control for gender and age, although these variables are not represent in Fig. 1 in order to improve readability. Additionally, as suggested by Preacher and Hayes (2008) for path analysis models including parallel multiple mediators, we let the residual terms for the SS-N and SS-NeR variables to covary, but this path is not shown in Fig. 1. In this context, fixing the covariance between SS-N and SS-NeR to zero is theoretically unreasonable as it would imply that the covariance existing between these mediators should only explained by their common predictors.

Because we test a saturated model which is expected to yield perfect fit, model fit was evaluated on the trimmed model (i.e., the model in which non-significant paths for null effects were constrained to zero). In establishing model fit, we use the comparative fit (CFI), the Tucker-

Table 1
Correlation between study measures (N = 472).

		M/%	SD	1	2	3	4	5	6	7
1	Age	13.49	1.07							
2	Gender (Male = 1, Female = 0)	49.60%	–	0.07						
3	SS – N	7.69	3.79	−0.03	−0.07	**0.80**				
4	SS – NeR	9.98	4.53	−0.09*	−0.09*	0.39**	**0.79**			
5	FoMO	22.56	7.09	0.05	−0.17**	0.31**	0.26**	**0.85**		
6	SMA	12.72	4.79	0.07	−0.12**	0.33**	0.14**	0.49**	**0.73**	
7	ES	8.73	2.55	0.07	−0.33**	0.11*	0.16**	0.40**	0.22**	**0.76**

Note. *p < .05, **p < .01. Values in bold represent Cronbach's alpha reliability coefficients. FoMO: Fear of Missing Out; SS-N: Sensitivity to Stress associated with Neglect; SS-NeR: Sensitivity to Stress associated with Negative Reactions; SMA: Social Media Addiction.

Lewis (TLI) and the root mean-square error of approximation (RMSEA) indexes. We consider values of CFI > 0.95, TLI > 0.95 and RMSEA < 0.05 as indication of good model fit, while CFI and TLI values ≤ 0.95 but > 0.90, and RMSEA ≥ 0.05 but < 0.08, indicate acceptable fit (Hu & Bentler, 1999). Analyses were performed using MPLUS, version 8.

3. Results

Results of reliability analyses, descriptive statistics, and correlations among study measures are shown in Table 1. All administered instruments showed adequate reliability (α ≥ 0.70). As regards correlations, a weak negative correlation emerged between age and SS-NeR on social media. In turn, being male was negatively correlated with SS-NeR, FoMO, SMA, and ES. SS-N and SS-NeR showed a positive moderate inter-correlation and a similar pattern of positive correlations with FoMO; SMA, and ES. Finally, FoMO was positively correlated with SMA and ES, which also showed a positive inter-correlation.

Next, regression-based path analysis showed the existence of many direct and indirect effects. Results of the estimation of direct path coefficients for the saturated model are shown in Fig. 1, while estimated indirect effects are reported in Table 2. FoMO emerged as a direct positive predictor of SS-N and SS-NeR, SMA, and ES. SS-N, but not SS-NeR, emerged as a positive predictor of SMA, which in turn showed a direct effect on ES. As regards indirect effects linking FoMO to ES, findings supported the existence of both a simple mediation effect passing through SMA and a serial mediation effect passing through both S-N and SMA (see Table 2). Concerning control variables, age showed a negative effect on SS-NeR ($B = -0.24$, 95% CI [−0.44, −0.04], $\beta = -0.10$), while gender (being male) showed a negative effect on ES ($B = -1.34$, 95% CI [−1.78, −0.97], $\beta = -0.27$). No remaining significant effects emerged.

Finally, we examined the fit of the model after removing non-significant paths for main effects. More specifically, we trimmed the tested model by fixing non-significant paths to zero (i.e., dashed lines in Fig. 1). The trimmed model showed excellent fit (χ^2 (2) = 3.090, p = .213; CFI = 0.997; TLI = 0.997; $RMSEA$ = 0.034).

Table 2
Fear of missing out predicting emotional symptoms: Estimates and confidence interval of indirect effects via fear of neglect and negative reactions, and social media addiction.

Paths of indirect effects	B	β	LLCI	ULCI
FoMO → SS-N → ES	−0.008	−0.023	−0.020	0.002
FoMO → SS-NeR → ES	0.007	0.020	−0.003	0.016
FoMO → SMA → ES	0.025	0.070	0.010	0.040
FoMO → SS-N → SMA → ES	0.004	0.012	0.001	0.008
FoMO → SS-NeR → SMA → ES	−0.001	−0.002	−0.003	0.001

Note. LLCI: Lower Level Confidence Interval; ULCI: Upper Level Confidence Interval. FoMO: Fear of Missing Out; SS-N: Sensitivity to Stress associated with Neglect; SS-NeR: Sensitivity to Stress associated with Negative Reactions; SMA: Social Media Addiction.

4. Discussion

The main aim of the present study was to investigate the interplay between FoMO and social media addiction in explaining individual differences in emotional symptoms in adolescence. We followed the hypothesis that FoMO could be associated with increased emotional symptoms in adolescents by route of heightened risk of social media addiction. Further, we looked at possible intervening factors, exploring the hypothesis that a heightened sensitivity to stress associated with negative experiences on social media when interacting with online peers (Beyens et al., 2016) might contribute to explain the links between FoMO, increased risk for social media addiction, and negative emotionality in adolescence.

In keeping with previous findings (Baker et al., 2016; Milyavskaya et al., 2018), our study supported the relationship between FoMO and decreased emotional well-being in adolescents. Next, we found FoMO was associated with higher sensitivity to stress associated with experiences of to neglect and negative reactions by online peers on social media. Several theorists have pointed out that subjects with high levels of FoMO tend to have higher group membership needs and/or popularity, and that FoMO mediates the relationship between these needs and distress (Beyens et al., 2016; Oberst et al., 2017). As such, it reasonable to expect that FoMO might increase perceptions of stress due to neglect or negative reactions on social media (Beyens et al., 2016), and overall negative emotionality (Valkenburg et al., 2006). Further, in line with previous literature (Al-Menayes, 2016; Blackwell et al., 2017; Dhir et al., 2018; Franchina et al., 2018), our findings support the link between FoMO and SMA, and identifies in an heightened sensitivity to stress associated with neglect (but not with negative reactions) by online peers a possible factor mediating their relationship. Thus, FoMO appears to fuel adolescents' stress related to not receiving feedback from their immediate online social network, which in turn appears to trigger addiction to social media platforms. This dynamic seems to agree with the I-PACE model, originally designed to explain Internet addiction, and recently extended to other forms of addiction behavior. Based on this complex model, we can interpret FoMO as a specific need and a predisposing characteristic of addiction in adolescents, which activates a state of distress when the subject assesses the situation as unsatisfactory for his/her own needs. The excessive use of social media and the development of SMA can therefore be seen as affective and cognitive responses aimed at restoring gratification or compensation with respect to perceived needs. In accordance with Wegmann and Brand (2019), FoMO can influence the development of SMA through two hypothetical mechanisms: the first based on a fear-driven / compensation-seeking hypothesis and the second based on a reward-driven hypothesis. In the fear-driven / compensation-seeking hypothesis, FoMO activates typical SMA behaviors by determining negative reinforcement aimed at reducing the fear of isolation and FoMO and the heightened stress deriving from them. In the reward-driven hypothesis, the use of social media would generate positive reinforcement leading to an increase in gratification and the satisfaction of social needs.

In this view, spending increasing amount of time on social media is

likely to represent a cognitive-emotional regulation strategy aimed at managing the stress associated with failing to fulfill psychological needs of belonging and popularity among online peers (e.g., failing to engage online peers), which in turn are fueled by high FoMO levels. However, such strategy can be dysfunctional, leading to addictive social media behaviors, and ultimately, an increase in emotional symptoms.

A novel, interesting finding emerging from the present study relates to lack of association between sensitivity to stress associated with negative reactions by online peers (e.g., receiving negative comments, or being kicked out of a social media group), social media addiction, and emotional symptoms, when controlling for sensitivity to stress related to neglect by online peers (e.g., not receiving Likes, or comments). Although no previous findings existing investigating these components adolescents' social media experience as separate facets, as well as their relations with psychological outcomes, our findings appear to be in line with reported by Zell and Moeller (2018) concerning the differential impact of receiving Likes, and experiencing negative comments to their posts, on the perception of online users about the importance of their posts. Among young adult social media users, not receiving Likes turned out to be more impactful than receiving either negative or positive comments in establishing the importance of their online posting activity; it is reasonable to expect, a similar differential effect might also apply to adolescent social media users rating their social status based on social media metrics.

In spite of the theoretical interpretability, and the empirical support existing for the relationship between social media addiction and emotional symptoms (Bányai et al., 2017; Dhir et al., 2018; Kircaburun et al., 2019; Pontes, 2017; Raudsepp & Kais, 2019; Settanni et al., 2018; Worsley et al., 2018), it is important to note that causal relationship between the two constructs is not entirely clear and it is possible that they present a two-way relationship. However, our study being cross-sectional in nature represents an important limitation, compromising our ability to provide a precise answer on the direction of causality of both this and other relationships emerging from our findings, that, within the framework of the model presented here, should be further investigated using longitudinal approaches.

5. Conclusion

Stress related to life on social media, and in particular the fear of being neglected by one's online social network, seems to fuel dependence on social media platforms. This type of stress would therefore seem to be reflected in a problematic or excessive use of social media, possibly aimed at maintaining or increasing one's online relationships, satisfying adolescents' need to belong to a peer group. However, this appears to be a dysfunctional strategy, as it may lead to a worsening in adolescents' psychological well-being.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.addbeh.2020.106364.

References

Al-Menayes, J. (2016). The fear of missing out scale: Validation of the Arabic version and correlation with social media addiction. *International Journal of Applied Psychology, 6*(2), 41–46. https://doi.org/10.5923/j.ijap.20160602.04.

Andreassen, C. S., Torsheim, T., Brunborg, G. S., & Pallesen, S. (2012). Development of a Facebook Addiction Scale. *Psychological Reports, 110*, 501–517. https://doi.org/10.2466/02.09.18.PR0.110.2.501-517.

Andreassen, C. S., Billieux, J., Griffiths, M. D., Kuss, D. J., Demetrovics, Z., Mazzoni, E., & Pallesen, S. (2016). The relationship between addictive use of social media and video games and symptoms of psychiatric disorders: A large-scale cross-sectional study. *Psychology of Addictive Behaviors, 30*(2), 252.

Andreassen, C. S., Pallesen, S., & Griffiths, M. D. (2017). The relationship between addictive use of social media, narcissism, and self-esteem: Findings from a large national survey. *Addictive Behaviors, 64*, 287–293. https://doi.org/10.1016/j.addbeh.2016.03.006.

Badenes-Ribera, L., Fabris, M. A., Gastaldi, F. G. M., Prino, L. E., & Longobardi, C. (2019). Parent and peer attachment as predictors of facebook addiction symptoms in different developmental stages (early adolescents and adolescents). *Addictive Behaviors, 95*, 226–232. https://doi.org/10.1016/j.addbeh.2019.05.009.

Baker, Z. G., Krieger, H., & LeRoy, A. S. (2016). Fear of missing out: Relationships with depression, mindfulness, and physical symptoms. *Translational Issues in Psychological Science, 2*(3), 275.

Bányai, F., Zsila, Á., Király, O., Maraz, A., Elekes, Z., Griffiths, M. D., ... Demetrovics, Z. (2017). Problematic social media use: Results from a large-scale nationally representative adolescent sample. *PLoS One, 12*(1), e0169839. https://doi.org/10.1371/journal.pone.0169839.

Barker, V. (2009). Older adolescents' motivations for social network site use: The influence of gender, group identity, and collective self-esteem. *Cyberpsychology & Behavior, 12*(2), 209–213. https://doi.org/10.1089/cpb.2008.0228.

Beyens, I., Frison, E., & Eggermont, S. (2016). "I don't want to miss a thing": Adolescents' fear of missing out and its relationship to adolescents' social needs, Facebook use, and Facebook related stress. *Computers in Human Behavior, 64*, 1–8. https://doi.org/10.1016/j.chb.2016.05.083.

Blackwell, D., Leaman, C., Tramposch, R., Osborne, C., & Liss, M. (2017). Extraversion, neuroticism, attachment style and fear of missing out as predictors of social media use and addiction. *Personality and Individual Differences, 116*, 69–72. https://doi.org/10.1016/j.paid.2017.04.039.

Brand, M., Wegmann, E., Stark, R., Müller, A., Wölfling, K., Robbins, T. W., & Potenza, M. N. (2019). The Interaction of Person-Affect-Cognition-Execution (I-PACE) model for Addictive Behaviors: Update, generalization to Addictive Behaviors beyond Internet-use disorders, and specification of the process character of Addictive Behaviors. *Neuroscience & Biobehavioral Reviews*. https://doi.org/10.1016/j.neubiorev.2019.06.032.

Brand, M., Young, K. S., Laier, C., Wölfling, K., & Potenza, M. N. (2016). Integrating psychological and neurobiological considerations regarding the development and maintenance of specific Internet-use disorders: An Interaction of Person-Affect-Cognition-Execution (I-PACE) model. *Neuroscience & Biobehavioral Reviews, 71*, 252–266. https://doi.org/10.1016/j.neubiorev.2016.08.033.

Casale, S., Caplan, S. E., & Fioravanti, G. (2016). Positive metacognitions about Internet use: The mediating role in the relationship between emotional dysregulation and problematic use. *Addictive Behaviors, 59*, 84–88. https://doi.org/10.1016/j.addbeh.2016.03.014.

Casale, S., Rugai, L., & Fioravanti, G. (2018). Exploring the role of positive metacognitions in explaining the association between the fear of missing out and social media addiction. *Addictive Behaviors, 85*, 83–87. https://doi.org/10.1016/j.addbeh.2018.05.020.

Cerniglia, L., Griffiths, M. D., Cimino, S., De Palo, V., Monacis, L., Sinatra, M., & Tambelli, R. (2019). A latent profile approach for the study of internet gaming disorder, social media addiction, and psychopathology in a normative sample of adolescents. *Psychology research and behavior management, 12*, 651. https://doi.org/10.2147/PRBM.S211873.

Cerniglia, L., Guicciardi, M., Sinatra, M., Monacis, L., Simonelli, A., & Cimino, S. (2019). The Use of Digital Technologies, Impulsivity and Psychopathological Symptoms in Adolescence. *Behavioral Sciences, 9*(8), 82. https://doi.org/10.3390/bs9080082.

Ciarrochi, J., Parker, P., Sahdra, B., Marshall, S., Jackson, C., Gloster, A. T., & Heaven, P. (2016). The development of compulsive internet use and mental health: A four-year study of adolescence. *Developmental Psychology, 52*(2), 272.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum.

Dhir, A., Yossatorn, Y., Kaur, P., & Chen, S. (2018). Online social media fatigue and psychological wellbeing-A study of compulsive use, fear of missing out, fatigue, anxiety and depression. *International Journal of Information Management, 40*, 141–152. https://doi.org/10.1016/j.ijinfomgt.2018.01.012.

Franchina, V., Vanden Abeele, M., van Rooij, A., Lo Coco, G., & De Marez, L. (2018). Fear of missing out as a predictor of problematic social media use and phubbing behavior among Flemish adolescents. *International Journal of Environmental Research and Public Health, 15*(10), 2319. https://doi.org/10.3390/ijerph15102319.

Goodman, R., Meltzer, H., & Bailey, V. (1998). The Strengths and Difficulties

Questionnaire: A pilot study on the validity of the self-report version. European child & adolescent psychiatry, 7(3), 125–130.
Griffiths, M. (2005). A 'components' model of addiction within a biopsychosocial framework. Journal of Substance use, 10(4), 191–197. https://doi.org/10.1080/14659890500114359.
Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal, 6(1), 1–55. https://doi.org/10.1080/10705519909540118.
Kemp, S. (2017). Digital in 2017: Global overview. We are social, 24.
Kircaburun, K., Kokkinos, C. M., Demetrovics, Z., Király, O., Griffiths, M. D., & Çolak, T. S. (2019). Problematic online behaviors among adolescents and emerging adults: Associations between cyberbullying perpetration, problematic social media use, and psychosocial factors. 1-908 International Journal of Mental Health and Addiction, 17(4), 891–908. https://doi.org/10.1007/s11469-018-9894-8.
Li, J. B., Mo, P. K., Lau, J. T., Su, X. F., Zhang, X., Wu, A. M., ... Chen, Y. X. (2018). Online social networking addiction and depression: The results from a large-scale prospective cohort study in Chinese adolescents. Journal of Behavioral Addictions, 7(3), 686–696. https://doi.org/10.1556/2006.7.2018.69.
Marengo, D., Longobardi, C., Fabris, M. A., & Settanni, M. (2018). Highly-visual social media and internalizing symptoms in adolescence: The mediating role of body image concerns. Computers in Human Behavior, 82, 63–69. https://doi.org/10.1016/j.chb.2018.01.003.
Marengo, D., Poletti, I., & Settanni, M. (2020). The interplay between neuroticism, extraversion, and social media addiction in young adult Facebook users: Testing the mediating role of online activity using objective data. Addictive Behaviors, 102, 106150. https://doi.org/10.1016/j.addbeh.2019.106150.
Marzocchi, G. M., Capron, C., Di Pietro, M., Tauleria, E. D., Duyme, M., Frigerio, A., ... Thérond, C. (2004). The use of the Strengths and Difficulties Questionnaire (SDQ) in Southern European countries. European Child & Adolescent Psychiatry, 13(2), ii40–ii46. https://doi.org/10.1007/s00787-004-2007-1.
Milyavskaya, M., Saffran, M., Hope, N., & Koestner, R. (2018). Fear of missing out: Prevalence, dynamics, and consequences of experiencing FOMO. Motivation and Emotion, 42(5), 725–737. https://doi.org/10.1007/s11031-018-9683-5.
Monacis, L., De Palo, V., Griffiths, M. D., & Sinatra, M. (2017). Social networking addiction, attachment style, and validation of the Italian version of the Bergen Social Media Addiction Scale. Journal of Behavioral Addictions, 6(2), 178–186. https://doi.org/10.1556/2006.6.2017.023.
Monacis, L., Palo, V. D., Griffiths, M. D., & Sinatra, M. (2016). Validation of the internet gaming disorder scale–short-form (IGDS9-SF) in an Italian-speaking sample. Journal of Behavioral Addictions, 5(4), 683–690. https://doi.org/10.1556/2006.5.2016.083.
Montag, C., Wegmann, E., Sariyska, R., Demetrovics, Z., & Brand, M. (2019). How to overcome taxonomical problems in the study of Internet use disorders and what to do with "smartphone addiction"? Journal of Behavioral Addictions, 1–7. https://doi.org/10.1556/2006.8.2019.59.
Nesi, J., Choukas-Bradley, S., & Prinstein, M. J. (2018). Transformation of adolescent peer relations in the social media context: Part 1-a theoretical framework and application to dyadic peer relationships. Clinical Child and Family Psychology Review, 21(3), 267–294. https://doi.org/10.1007/s10567-018-0261-x.
Oberst, U., Wegmann, E., Stodt, B., Brand, M., & Chamarro, A. (2017). Negative consequences from heavy social networking in adolescents: The mediating role of fear of missing out. Journal of Adolescence, 55, 51–60. https://doi.org/10.1016/j.adolescence.2016.12.008.
Pontes, H. M. (2017). Investigating the differential effects of social networking site addiction and Internet gaming disorder on psychological health. Journal of Behavioral addictions, 6(4), 601–610. https://doi.org/10.1556/2006.6.2017.075.
Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. Behavior Research Methods, 40(3), 879–891. https://doi.org/10.3758/BRM.40.3.879.
Przybylski, A. K., Murayama, K., DeHaan, C. R., & Gladwell, V. (2013). Motivational, emotional, and behavioral correlates of fear of missing out. Computers in Human Behavior, 29(4), 1841–1848. https://doi.org/10.1016/j.chb.2013.02.014.
Raudsepp, L., & Kais, K. (2019). Longitudinal associations between problematic social media use and depressive symptoms in adolescent girls. Preventive Medicine Reports, 100925. https://doi.org/10.1016/j.pmedr.2019.100925.
Rideout, V., & Robb, M. B. (2018). Social media, social life: Teens reveal their experiences. San Francisco, CA: Common Sense Media.
Settanni, M., Marengo, D., Fabris, M. A., & Longobardi, C. (2018). The interplay between ADHD symptoms and time perspective in addictive social media use: A study on adolescent Facebook users. Children and Youth Services Review, 89, 165–170. https://doi.org/10.1016/j.childyouth.2018.04.031.
Valkenburg, P. M., Peter, J., & Schouten, A. P. (2006). Friend networking sites and their relationship to adolescents' well-being and social self-esteem. CyberPsychology & Behavior, 9(5), 584–590.
Van Den Eijnden, R. J., Lemmens, J. S., & Valkenburg, P. M. (2016). The social media disorder scale. Computers in Human Behavior, 61, 478–487. https://doi.org/10.1016/j.chb.2016.03.038.
Wegmann, E., & Brand, M. (2019). A narrative Overview About Psychosocial Characteristics as Risk Factors of a Problematic Social Networks Use. Current Addiction Reports, 6(4), 402–409. https://doi.org/10.1007/s40429-019-00286-8.
Worsley, J. D., McIntyre, J. C., Bentall, R. P., & Corcoran, R. (2018). Childhood maltreatment and problematic social media use: The role of attachment and depression. Psychiatry Research, 267, 88–93. https://doi.org/10.1016/j.psychres.2018.05.023.
Zell, A. L., ... Moeller, L. (2018). Are you happy for me... on Facebook? The potential importance of "likes" and comments. Computers in Human Behavior, 78, 26–33. https://doi.org/10.1016/j.chb.2017.08.050.

6

Telematics and Informatics 101 (2025) 102309

Contents lists available at ScienceDirect

Telematics and Informatics

journal homepage: www.elsevier.com/locate/tele

For you vs. for everyone: The effectiveness of algorithmic personalization in driving social media engagement

Cynthia A. Dekker *, Susanne E. Baumgartner, Sindy R. Sumter

Amsterdam School of Communication Research, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, the Netherlands

Algorithm / Feelings

T

atforms increasingly use algorithmic personalization, raising concerns about po- lled usage. However, these concerns remain partly speculative as evidence for the algorithmic personalization in driving user engagement is limited. Therefore, the vestigated how TikTok users' behavior and experiences would change if their onger personalized based on their interests. In this preregistered study, 88 TikTok ed in a two-week within-subjects design: a baseline week (default highly ed), followed by an experimental week (less personalized feed). Daily experiences hrough daily surveys, and objective TikTok usage data was obtained through e found that both daily frequency and duration of TikTok use decreased, self- eased, and participants derived less enjoyment from their use. These findings ritical role of algorithmic personalization in sustaining user engagement and educing feed personalization may be a promising, though currently limited, dress uncontrolled social media use.

r et al., 2022; Vanden Abeele & Nguyen, 2024; Ytre-Arne et al., 2020). On ctivity and entertainment experiences these platforms provide, yet on the ir use. Media psychological research increasingly points to platform design features as key contributors to such ambivalent social media experiences (e.g., Flayelle et al., 2023; Vanden Abeele, 2021). A prominent example is the algorithmically personalized feed, which presents users with content tailored to their inferred interests. Although this personalization can enhance the user experience by making it maximally entertaining, it may also lead to uncontrolled use as users find it difficult to stop scrolling (e.g., Ilg et al., 2023).

TikTok stands out in this regard. Its algorithmic personalization outperforms other social media platforms according to its users (Bhandari and Bimo 2022; Lee et al., 2022; Schellewald, 2023; Taylor & Choi, 2022), and it is seen as the main reason for the short-format video app's popularity (e.g., Bucker & Urman, 2022; Feldkamp, 2021). However, alongside the praise for TikTok's accuracy in curating content that appeals to users' interests, concerns have been raised about potential negative consequences of algorithmic personalization. For example, users report spending more time on the app than intended and procrastinating other tasks (Kang & Lou, 2022; Ramsden & Talbot, 2024). Both users and politicians have thus criticized the "addictive nature" of the personalized feed (Allyn,

* Corresponding author.
E-mail address: C.A.Dekker@uva.nl (C.A. Dekker).

https://doi.org/10.1016/j.tele.2025.102309
Received 17 December 2024; Received in revised form 28 May 2025; Accepted 6 June 2025
Available online 8 June 2025
0736-5853/

Telematics and Informatics 101 (2025) 102300

Contents lists available at ScienceDirect

Telematics and Informatics

journal homepage: www.elsevier.com/locate/tele

For you vs. for everyone: The effectiveness of algorithmic personalization in driving social media engagement

Cynthia A. Dekker *, Susanne E. Baumgartner, Sindy R. Sumter

Amsterdam School of Communication Research, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, the Netherlands

ARTICLE INFO

Keywords:
Social media
Algorithmic personalization
Algorithmic feed
TikTok
Enjoyment
Entertainment
Self-regulation

ABSTRACT

Social media platforms increasingly use algorithmic personalization, raising concerns about potential uncontrolled usage. However, these concerns remain partly speculative as evidence for the effectiveness of algorithmic personalization in driving user engagement is limited. Therefore, the present study investigated how TikTok users' behavior and experiences would change if their feeds were no longer personalized based on their interests. In this preregistered study, 88 TikTok users participated in a two-week within-subjects design: a baseline week (default highly personalized feed), followed by an experimental week (less personalized feed). Daily experiences were assessed through daily surveys, and objective TikTok usage data was obtained through screenshots. We found that both daily frequency and duration of TikTok use decreased, self-regulation increased, and participants derived less enjoyment from their use. These findings highlight the critical role of algorithmic personalization in sustaining user engagement and suggest that reducing feed personalization may be a promising, though currently limited, approach to address uncontrolled social media use.

1. Introduction

Social media use is often experienced ambivalently (Faber et al., 2022; Vanden Abeele & Nguyen, 2024; Ytre-Arne et al., 2020). On the one hand, users appreciate the always-available connectivity and entertainment experiences these platforms provide, yet on the other hand, they also frequently struggle with regulating their use. Media psychological research increasingly points to platform design features as key contributors to such ambivalent social media experiences (e.g., Flayelle et al., 2023; Vanden Abeele, 2021). A prominent example is the algorithmically personalized feed, which presents users with content tailored to their inferred interests. Although this personalization can enhance the user experience by making it maximally entertaining, it may also lead to uncontrolled use as users find it difficult to stop scrolling (e.g., Ye et al., 2023).

TikTok stands out in this regard. Its algorithmic personalization outperforms other social media platforms according to its users (Bhandari and Bimo 2022; Lee et al., 2022; Schellewald, 2023; Taylor & Choi, 2022), and it is seen as the main reason for the short-format video app's popularity (e.g., Boeker & Urman, 2022; Feldkamp, 2021). However, alongside the praise for TikTok's accuracy in curating content that appeals to users' interests, concerns have been raised about potential negative consequences of algorithmic personalization. For example, users report spending more time on the app than intended and procrastinating other tasks (Kang & Lou, 2022; Ramsden & Talbot, 2024). Both users and politicians have thus criticized the "addictive nature" of the personalized feed (Allyn,

* Corresponding author.
E-mail address: C.A.Dekker@uva.nl (C.A. Dekker).

https://doi.org/10.1016/j.tele.2025.102300
Received 17 December 2024; Received in revised form 28 May 2025; Accepted 6 June 2025
Available online 8 June 2025

2024; Chen et al., 2023; European Commission, 2024; Kang & Lou, 2022).

While prior research has begun to illuminate how algorithmically personalized feeds shape user experiences and engagement, much of this evidence has come from qualitative or cross-sectional studies. These approaches have provided important insights, such as that higher perceived accuracy of algorithmic personalization is associated with greater enjoyment (e.g., Taylor & Choi, 2022). Yet, experimental studies that manipulate algorithmic personalization in real-world contexts remain rare, partly due to the methodological challenges involved in such designs. One notable exception is a study showing that switching from an algorithmically personalized feed to a chronologically ordered feed on Facebook and Instagram reduced time spent on these platforms (Guess et al., 2023). Still, little is known about how TikTok users respond when the personalized feed becomes less personalized: What changes in their usage, entertainment experiences, self-regulation, and mental preoccupation with the app? Such insights are important for understanding the impact of algorithmic personalization in shaping user engagement and may inform future legislative efforts aimed at reducing potential harms for users.

The current study seeks to contribute to this emerging field by leveraging the newly introduced setting that TikTok implemented in compliance with the European Union's Digital Services Act (DSA; European Commission, 2024). Specifically, TikTok now allows European users to "switch off" the personalization of their feeds. When using this less personalized version of the feed, users see "locally relevant and globally 'popular' videos" (TikTok, 2023) instead of content based on algorithmic inferences about their interests. The setting thus offers the opportunity to experimentally assess how users' TikTok usage and experiences change when they switch from a highly personalized to a less personalized feed.

To this end, we conducted a quasi-experimental within-subjects field study in which TikTok users switched to a less personalized feed for one week, with daily measures of subjective user experiences (entertainment experiences; self-regulation; TikTok vigilance) as well as objective usage metrics (time spent; use frequency).

1.1. *The TikTok algorithm*

Unlike other social media platforms where networks between users play a key role in content distribution (e.g., friends, connections, followers), TikTok is built around its algorithm-based content recommendations feed called the For You page (Bhandari and Bimo 2022; Boeker & Urman, 2022; Lee et al., 2022). The For You page is the main feature and landing page of the app, where users can watch an infinite number of videos selected for them by TikTok. When opening the app, the first recommended video automatically starts playing, and users can scroll to see the next video. While TikTok offers common social media features that allow users to express an explicit preference for certain content (e.g., following other users; liking or sharing videos), this is not essential for their stream of personalized content as the algorithm learns about users' interests from a variety of factors. According to TikTok, these factors include engagement metrics (e.g., viewing time), metrics related to the videos one engaged with (e.g., hashtags), and device and account settings (e.g., language; TikTok, 2020). Based on these factors as well as data from similar users, the algorithm can predict the likelihood of a user's interest in a certain video. The more frequently and longer someone uses TikTok—producing more data—the more accurate these inferences will be (TikTok, 2020; Zhao, 2021).

However, the exact workings of TikTok's algorithmic system remain opaque (Boeker & Urman, 2022; Zhao, 2021). That is, the precise selection of factors informing the algorithm and their relative weight in prioritizing and filtering content are largely unkown. This lack of transparency is further complicated by the proprietary and constantly evolving nature of the model, making it difficult to determine precisely why certain videos are shown to users (e.g., Baumann et al., 2025).

Several studies have attempted to lift this veil by, for instance, manually interacting (Hilbert et al., 2023) or automatizing TikTok accounts through bots with varying user interaction patterns (Boeker & Urman, 2022; Vombatkere et al., 2024), exploring metrics of trending videos (Klug et al., 2021), or examining the invention patents related to TikTok's algorithm (Zhao, 2021). In general, they found that liking, following, and viewing time are most predictive of content personalization (e.g., Boeker & Urman, 2022). Further, it has been found that in addition to simply catering to users' known interests (i.e., interest exploitation), the TikTok feed sometimes delivers videos that do not perfectly match the inferred interests but may reveal new interests (i.e., interest exploration; Vombatkere et al., 2024). This unpredictable content delivery pattern ensures the algorithm can keep up with users' naturally evolving interests (Lee et al., 2022) and prevents users from getting bored by too much repetition (Anderson & Wood, 2020; Zhao, 2021).

Another line of research focuses on users' awareness and perspectives of social media algorithms. Users are generally aware of their existence and have formed beliefs about how they function, such as how their interactions with the algorithm are reciprocal, as the algorithm adapts based on the user's interactions with the presented content (Bucher, 2017; Lee et al., 2022; Schellewald, 2022). Users also attempt to actively influence the algorithm by strategically interacting with content in certain ways (Bucher, 2017; Lee et al., 2022; Schellewald, 2022). Consequently, they feel like, over time, the algorithm gets to know their multifaceted, dynamic interests and even reflects their identity (Bhandari and Bimo 2022; Lee et al., 2022).

For the present study, then, we need to acknowledge that the technical workings behind TikTok's personalized feed partially remain a black box. Notably, this also applies to the feed when personalization is disabled. This "non-personalized" feed is still not completely random, as the content is curated based on popularity and the user's region, thus still personalized to some extent (TikTok, 2023). We, therefore, label the default personalized TikTok feed as the *highly personalized* feed and the personalization-disabled feed as the *less personalized feed*. The key difference is that the highly personalized feed presents content based on users' inferred interests, while the less personalized feed does not. Thus, despite the uncertainties around what these feeds technically entail, and their evolving nature, our study sheds light on how this aspect of algorithmic personalization (i.e., personalization based on inferred user interests) impacts user engagement and experiences.

1.2. *Time spent and use frequency*

Algorithmic personalization has been linked to increased usage (e.g., Montag & Hegelich, 2020). Adult users spend about one hour per day on TikTok, which surpasses the time spent on other social media platforms (De Marez et al., 2024; Feger, 2024). Young adults aged 18–24 even spend around 80 min per day on TikTok on average (De Marez et al., 2024; Feger, 2024). When it comes to frequency of usage, TikTok reported that its average user opens the app 19 times per day (Dua, 2021). With the personalization of the feed making TikTok "both appealing and hard to leave once there" (Schellewald, 2023, p. 1574), it seems likely that people would reduce their usage if their feed is less personalized (Montag & Hegelich, 2020). In line with this reasoning, an experimental study showed that switching from an algorithmically personalized feed to a chronological feed reduced time spent on Facebook and Instagram (Guess et al., 2023). We thus predicted that:

H1: Both (a) daily time spent on TikTok and (b) daily frequency of opening TikTok will be significantly lower when using a less personalized feed compared to a highly personalized feed.

1.3. *Entertainment experiences*

In studies examining motivations for using TikTok, entertainment was consistently identified as a key motive (e.g., Bucknell Bossen & Kottasz, 2020; Dong & Xie, 2024; Lu et al., 2020; Scherr & Wang, 2021; Vaterlaus & Winter, 2021). Therefore, we investigate how the level of feed personalization relates to entertainment experiences on the platform. Personal relevance of content plays a crucial role in shaping such experiences (Vorderer et al., 2004), and entertainment experiences typically manifest in two ways: enjoyment and meaningfulness (Oliver & Bartsch, 2010). Moreover, when an entertainment experience is particularly engaging, users can lose track of time (Sherry, 2004). Accordingly, we focus on four outcome variables: personal relevance, enjoyment, meaningfulness, and time distortion.

Personal relevance of content, defined as the perceived alignment with one's interests (Vorderer et al., 2004), may be a direct result of high personalization. Qualitative studies show that users believe TikTok's highly personalized feed to be remarkably accurate in providing personally relevant content (Bhandari and Bimo 2022; Kang & Lou, 2022; Lee et al., 2022; Martinez et al., 2024).

In addition, TikTok users commonly describe their experience with the highly personalized feed as enjoyable (e.g., Lee et al., 2022; Vaterlaus & Winter, 2021), defined as a pleasant response state (Vorderer et al., 2004). A survey study found that users' enjoyment was positively correlated with the perceived accuracy of TikTok's algorithmic system (Taylor & Choi, 2022). Besides enjoyment, users report deriving meaningful experiences from watching personalized TikTok content, marked by a deeper level of processing and appreciation (Oliver & Bartsch, 2010). For example, users describe how some videos serve as sources of inspiration and life lessons (Lu et al., 2020; Martinez et al., 2024) or provide a sense of belonging and relatability (Lee et al., 2022; Martinez et al., 2024). Feed personalization can thus evoke appreciation and introspection by recommending videos that go beyond mere fun (Lu et al., 2020).

Finally, due to an infinite stream of engaging content, users report they often are so engrossed in scrolling that they lose track of time, an experience referred to as time distortion (Kang & Lou, 2022; Martinez et al., 2024; Ramsden & Talbot, 2024; Schellewald, 2022; Vaterlaus & Winter, 2021). Correlational studies have found that time distortion is related to the relevance and serendipity of recommended TikTok videos as well as the ease of viewing them (i.e., not having to search for relevant content; Yang et al., 2023; Zhao & Wagner, 2022).

In sum, TikTok's highly personalized feed provides an entertainment experience characterized by personal relevance, enjoyment, meaningfulness, and time distortion. Since these experiences are likely the result of personally appealing content, we proposed that:

H2: Participants will report lower levels of (a) personal relevance, (b) TikTok use enjoyment, (c) meaningfulness of the TikTok content, and (d) time distortion while using TikTok, when using a less personalized feed compared to a highly personalized feed.

1.4. *Self-regulation of TikTok use*

Although users sometimes want to "get carried away by the algorithm" (Schellewald, 2023), they also frequently struggle with regulating their TikTok use. These experiences of low self-regulation are often attributed to algorithmic personalization (e.g., Montag & Hegelich, 2020). Therefore, we examine three indicators of TikTok use self-regulation, namely perceived control, automaticity of use, and procrastination.

When it comes to perceived control over one's usage, qualitative work has shown that people describe the highly personalized feed as "addictive", as they find it difficult to disengage once they start scrolling (Chen et al., 2023; Kang & Lou, 2022; Martinez et al., 2024; Ramsden & Talbot, 2024; Schellewald, 2022, 2023; Vaterlaus & Winter, 2021). The high personalization of the feed keeps users continuously curious about the next pleasantly surprising video (Kang & Lou, 2022; Ramsden & Talbot, 2024). As a result, they may continue scrolling for longer than they had intended. We conceptualise this as a perceived lack of control over one's TikTok use, reflecting users' subjective sense of being unable to regulate their use in line with their intentions.

In addition, users typically develop a certain automaticity in using TikTok, which refers to the tendency to engage in behavior without conscious intention (e.g., Anderson & Wood, 2020). Both opening the TikTok app and scrolling through one's personalized feed can become such routine behaviors (Dong & Xie, 2024; Wang & Scherr, 2022). This habituation is widely understood to result from initial and intermittent rewards associated with repeated behavior (e.g., Anderson & Wood, 2020; Bayer et al., 2022). TikTok users gain such rewards from the videos they see on their highly personalized feed: The content delivery system is appealing and unpredictable at the same time. Each scroll to a new video is effortless and yields an instant micro-reward, thereby reinforcing the behavior (Anderson & Wood, 2020; Bayer et al., 2022). Neurological research confirms this idea by showing that personalized TikTok

videos evoke higher activations in brain regions related to reward cognition than non-personalized videos (Su et al., 2021). In this study, we focus on the perceived automaticity of opening the app and scrolling through the feed, both of which contribute to the habitual nature of TikTok use.

Furthermore, with a platform that is so engaging, users often find themselves delaying other activities to continue scrolling. More specifically, users report using the app to procrastinate daily tasks such as studying and household chores (Holzmann, 2021; Ramadan & Talbot, 2024; Schellewald, 2023; Xie et al., 2023). Similarly, TikTok use may interfere with healthy sleep routines, as many people use the app in the late evening before going to bed or while in bed (Dang & Xie, 2024; Lu et al., 2020; Scherr & Wang, 2021). While mindless scrolling can help to unwind (Schellewald, 2023), prolonged late-night usage may delay sleeping, ultimately causing daytime fatigue (Lu et al., 2020; Wang & Scherr, 2022). We examine how these two forms of procrastination (i.e., procrastinating tasks or sleeping) are related to the level of feed personalization. We conceptualize task procrastination as using TikTok instead of getting an intended task done. Similarly, we conceptualize sleep procrastination as going to sleep later than intended because of using TikTok.

Given that high feed personalization has been linked to experiences of reduced control, habitual engagement, and procrastination, we expect that lowering the level of personalization will mitigate these effects.

H3: Participants will report (a) higher perceived control over their TikTok use, (b) lower automaticity in TikTok opening and (c) scrolling, and (d) lower task and (e) sleep procrastination, when using a less personalized feed compared to a highly personalized feed.

1.5. *TikTok vigilance*

Finally, people may develop a cognitive preoccupation with TikTok even when they are not actively engaged with it. Given that people use TikTok to stay up-to-date on peers, news, and trends (Bhandari and Bimo 2022), and have integrated the app into their daily routines (Schellewald, 2022; Wang & Scherr, 2022), it seems people are mentally engaged with TikTok beyond the moments they actually spend on the app. This cognitive preoccupation, termed vigilance, describes a state of awareness and anticipation toward the app and is conceptualized as comprising three subdimensions (Klimmt et al., 2017). To illustrate, people might mentally reference content they have encountered on TikTok (salience), they may be highly attentive to TikTok notifications (reactibility), and they may tend to continuously check the app (monitoring). This state of vigilance toward TikTok appears to stem from the perceived constant potential of encountering relevant content on the app, and may thus be reduced once the feed is less personalized. We predicted:

H4: Participants will report lower TikTok vigilance—(a) salience, (b) reactibility, and (c) monitoring—when using a less personalized feed compared to a highly personalized feed.

2. Method

2.1. *Procedure*

To test the hypotheses, we conducted a two-week within-subjects quasi-experimental study, in which participants were asked to use the less personalized TikTok feed for one week following a baseline week. The study (methodology, hypotheses, analysis plan) was preregistered on the Open Science Framework (OSF) and was approved by the Ethics Review Board of the authors' university. Participants were recruited through the university's online lab facilities. Participants had to be at least 18 years old, have had TikTok installed on their phones for at least four weeks, and use the default personalized feed. The latter two criteria ensured that participants were accustomed to using the highly personalized feed, which was presumed to be well-personalized after four weeks of usage (Baumann et al., 2025; Siles et al., 2024). Participants could choose to be compensated for their two-week participation with research credits or €15.

After providing informed consent, participants were instructed to install a research app (Avicenna) on their smartphones. Through the app, participants completed an entry survey including demographic questions and items about, for instance, the use of other social media apps. The next day (i.e., the first full day of participation) marked the first day of the baseline week. Every morning at 8:00, participants received a short survey with the key outcome measures regarding their TikTok use the previous day (e.g., enjoyment). At the end of every daily survey, participants were instructed to upload screenshots of their TikTok use metrics from the day before, which could be found in the TikTok app. After the baseline week, a longer survey was sent, including instructions to switch to the less personalized feed. In the following week, participants kept it that way and completed daily surveys as before. At the end of the second week, an exit survey was completed, including measures related to self-reported compliance and future intentions regarding feed personalization, as well as open questions about their experiences. On average, participants completed 13.48 daily surveys ($SD = 1.13$)

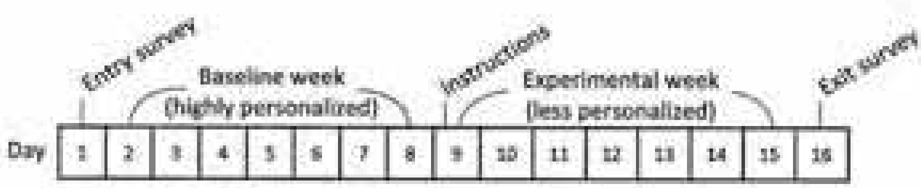


Fig. 1. *Schematic Representation of the Study Procedure. Note.* Daily surveys (with measures regarding the previous day) were sent in the morning on days 3 to 16. On day 9, the daily survey was followed by instructions to switch to the less personalized feed. On day 16, it was followed by the exit survey.

over the course of 14 days. A schematic overview of the study procedure can be found in Fig. 1.

2.2. *Sample*

Initially, 120 participants installed the research app and completed the entry survey. Applying the exclusion criteria, we excluded 18 participants for not completing the larger survey after the baseline week which included the instructions to switch to the less personalized feed, 3 participants for not completing the exit survey, 1 participant for not completing at least four daily surveys in both weeks, and an additional 10 participants for self-reported low compliance with the feed personalization instructions. The final sample consisted of $N = 88$ participants ($M_{age} = 21.97$, $SD_{age} = 3.08$), of which 85.2 % identified as woman and 14.8 % as man. Most participants (87.5 %) had been using TikTok for more than a year. In the entry survey, participants expressed dissatisfaction with how much time they spend on TikTok ("I feel like I spend too much time on TikTok"; 1 = *strongly disagree* to 7 = *strongly agree*; $M = 5.36$; $SD = 1.56$). They also wanted to reduce their TikTok use ("I want to reduce the time I spend on TikTok"; $M = 5.35$; $SD = 1.39$).[1]

2.3. *Measures*

All surveys and instructions can be found on OSF.

2.3.1. *TikTok use metrics*

Every TikTok user can find their use metrics in their TikTok app. Participants received detailed instructions on how to access these metrics and take correct screenshots (i.e., we provided screenshots of this procedure as well as example screenshots). Every day, participants uploaded two screenshots: one from their "time spent" metrics and one from their "TikTok opened" metrics (i.e., frequency of opening TikTok) of the day before. In the presentation of these metrics, TikTok distinguishes between day (6 AM – 10 PM) and night time (10 PM – 6 AM). The metrics displayed in the uploads from participants were manually registered into a dataset by the researcher. For the analyses, aggregated variables of TikTok use per day (i.e., 6 AM – 6 AM the next day) were used.

2.3.2. *Subjective experiences*

To ensure brevity of the daily surveys, and thus not overburden participants, single-item measurements were used to measure the subjective outcome variables. Importantly, recent work on the reliability and validity of single-item measures has shown that they can be appropriate assessments of unidimensional constructs (Martela & Ryan, 2024; Matthews et al., 2022; Wolters et al., 2023). All items referred to participants' experiences with TikTok the previous day. Response scales ranged from 1 (*strongly disagree*) to 7 (*strongly agree*), with some items including an additional response option (e.g., "I did not use TikTok yesterday) to ensure responses were valid even if the participant, for instance, had not used TikTok at all. These response options were marked as missing values for the analyses. An overview of all item wordings and references can be found in Table 1.

2.4. *Analytical strategy*

We tested the hypotheses using repeated measures analyses with a linear mixed models approach. Analyses were performed in R (v4.4.1; R Core Team, 2024), using lme4 (v1.1–35.5; Bates et al., 2015) and lmerTest (v3.1–3; Kuznetsova et al., 2017). In all models, week (0 = highly personalized feed vs. 1 = less personalized feed) was included as the predictor. In addition, age was included as a control variable, since small but significant correlations were found with some of the outcome variables (see Table 2). Following recommendations by Barr et al. (2013), all models were specified with a maximal random effects structure: a random intercept per participant and a random slope for the within-person predictor (i.e., week). For an overview of all means and standard deviations in both weeks, see Table 3. As fourteen outcomes were assessed, interpretations were based on a Bonferroni-corrected α-level of $p < 0.003$. The data and analysis script are available on OSF.

3. Results

3.1. *Confirmatory analyses*

We found significant decreases in both the average daily time spent on TikTok, $t(86.59) = -8.32$, $p < 0.001$, Cohen's $d = -0.78$, and the average daily frequency of opening TikTok, $t(89.05) = -6.67$, $p < 0.001$, $d = -0.49$. When using a less personalized feed, TikTok screen time was reduced by approximately 40 min per day, and app opening frequency was reduced by approximately five times per day, thus confirming H1a and H1b.

Participants reported that the TikTok videos they watched were significantly less personally relevant during the less personalized week, $t(88.06) = -14.86$, $p < 0.001$, $d = -2.46$, in line with H2a. They also found the content less enjoyable, $t(87.27) = -11.48$, $p < 0.001$, $d = -1.82$, and less meaningful, $t(88.36) = -9.83$, $p < 0.001$, $d = -1.46$, thus supporting H2b and H2c. In addition, participants

[1] We conducted an independent samples *t*-test to compare the final sample ($n = 88$) with participants who quit or were excluded ($n = 32$). The latter group reported slightly lower ($M = 4.78$, $SD = 1.86$), but not significantly lower, intentions to reduce their TikTok use, compared to those who did not drop out ($M = 5.35$, $SD = 1.39$), $t(118) = 1.81$, $p = 0.073$.

Table 1
Measures daily surveys.

Variable	Item	Adapted from
Entertainment experience		
Personal relevance	The TikTok content I viewed yesterday matched my interests.	Zhou & Wagner (2022)
Enjoyment	I enjoyed using TikTok yesterday.	–
Meaningfulness	The TikTok content I viewed yesterday was meaningful.	Oliver & Bartsch (2010)
Time distortion	I lost track of time while using TikTok yesterday.	Huang et al. (2022)
Self-regulation		
Perceived control	Yesterday, I used TikTok for longer than I had intended.[a]	Dekker & Baumgartner (2024)
Opening automaticity	Yesterday, opening the TikTok app was something I did automatically.	Gardner et al. (2012)
Scrolling automaticity	Yesterday, scrolling on my TikTok feed was something I did automatically.	Gardner et al. (2012)
Task procrastination	Yesterday, I used TikTok although I had planned to get something else done.	Meier et al. (2016)
Sleep procrastination	Yesterday, I went to sleep later than I had intended because of using TikTok.	Kroese et al. (2016)
TikTok vigilance		
Salience	Yesterday, my thoughts often drifted to TikTok, even when I was not using it.	Reinecke et al. (2018)
Reactibility	Yesterday, when I received a TikTok notification, it triggered an impulse in me to check TikTok right away.	Reinecke et al. (2018)
Monitoring	Yesterday, I often felt the urge to make sure I knew what was happening on TikTok.	Reinecke et al. (2018)

Note. Response scales ranged from 1 (*strongly disagree*) to 7 (*strongly agree*), with some items additionally including the option "I did not use TikTok yesterday" (personal relevance; enjoyment; meaningfulness; time distortion; perceived control) or "I did not receive a TikTok notification yesterday" (reactibility). [a] This item was reverse-coded: higher values indicate higher perceived control.

reported lower time distortion during TikTok use, $t(85.86) = -8.58$, $p < 0.001$, $d = -1.03$, supporting H2d.

Participants reported higher control over their TikTok use, $t(88.23) = 8.89$, $p < 0.001$, $d = 0.89$, when using a less personalized feed compared to the highly personalized feed, in line with H3a. Furthermore, supporting H3b and H3c, participants reported engaging less in automatic app opening, $t(89.04) = -6.37$, $p < 0.001$, $d = -0.59$, and automatic scrolling, $t(84.65) = -7.83$, $p < 0.001$, $d = -0.94$. Both task procrastination, $t(88.27) = -7.51$, $p < 0.001$, $d = -0.85$, and sleep procrastination, $t(89.19) = -6.38$, $p < 0.001$, $d = -0.61$, were also significantly reduced, supporting H3d and H3e.

For TikTok vigilance, we found significant reductions in both the salience and monitoring dimensions. Participants reported less thought-drifting toward TikTok (salience), $t(88.89) = -4.14$, $p < 0.001$, $d = -0.40$, and a lower tendency to monitor the app (monitoring), $t(86.20) = -3.94$, $p < 0.001$, $d = -0.35$. However, participants did not feel less reactive to notifications (reactibility), $t(58.56) = -1.33$, $p = 0.190$. Thus, H4a and H4c, but not H4b, were supported by our data.

3.2. *Additional analyses*

In addition to the confirmatory analyses, we explored whether participants had replaced their TikTok use with other activities. In the exit survey, we asked participants to what extent they perceived to have spent more or less time on other activities during the less personalized week than normally (1 = *much less than normally* to 7 = *much more than normally*). We ran one-sample *t*-tests to compare these mean scores to the midpoints of the scales. Participants reported spending somewhat more time on offline social activities (e.g., hanging out with friends or family), $M = 4.21$, $SD = 1.04$, $t(87) = 4.64$, $p < 0.001$, and other activities by themselves (e.g., reading, exercising, walking, or any other hobbies), $M = 4.44$, $SD = 0.96$, $t(87) = 4.34$, $p < 0.001$.

In addition, participants indicated that they had spent somewhat more time on other social media apps (e.g., Instagram, Snapchat, Facebook) than they normally do, $M = 4.68$, $SD = 1.38$, $t(87) = 4.64$, $p < 0.001$. This suggests that participants partially replaced their TikTok use with other social media. We were able to verify this perception, as participants were also asked to look up their average time spent on social media apps in their phone's screen time metrics (i.e., the "social media" category, including TikTok) for the week before the study ($M = 3{:}22$ h, $SD = 1{:}46$), the baseline week ($M = 3{:}40$, $SD = 2{:}49$), and the experimental week ($M = 2{:}47$, $SD = 1{:}43$). A paired samples *t*-test revealed that the reduction between the week before the study (i.e., "normally") and the experimental week was significant, $M_{\Delta} = 0.55$, $t(87) = 3.54$, $p < 0.001$, 95 % CI [0.26, 0.92]. The reduction between the baseline week and the experimental week was also significant, $M_{\Delta} = 0.53$, $t(87) = 3.37$, $p = 0.002$, 95 % CI [0.35, 1.43]. Given that participants' TikTok use decreased by 40 min in the experimental week compared to the baseline week, it appears that they did not compensate by spending more time on other social media apps, and in fact, slightly reduced their usage overall.

Finally, to gain a better understanding of participants' experience with the less personalized feed, we examined their future intentions. In the exit survey, we asked participants if they would immediately turn their highly personalized feeds back on (1 = *strongly disagree* to 7 = *strongly agree*). Most people agreed that they would return to their highly personalized feed, as only nine participants disagreed (i.e., response options 1–3) or responded neutrally (i.e., response option 4), $M = 5.90$, $SD = 1.37$.

4. Discussion

Recent political developments, such as the implementation of the DSA in Europe (European Commission, 2024) and lawsuits against TikTok in the US (Allyn, 2024), show that social media platforms are viewed with increasing scrutiny. A recurring theme is the

Table 2
Zero-order correlations baseline week.

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.
1. Time spent on TikTok	–															
2. TikTok use frequency	0.68***	–														
3. Personal relevance	0.24***	0.19***	–													
4. Enjoyment	0.06	0.02	0.52***	–												
5. Meaningfulness	0.06	0.03	0.35***	0.38***	–											
6. Time distortion	0.44***	0.32***	0.34***	0.18***	0.11**	–										
7. Perceived control	−0.49***	−0.36***	−0.33***	−0.17***	−0.08	−0.80***	–									
8. Opening automaticity	0.43***	0.43***	0.37***	0.24***	0.09*	0.56***	−0.56***	–								
9. Scrolling automaticity	0.39***	0.35***	0.50***	0.39***	0.12**	0.61***	−0.58***	0.79***	–							
10. Task procrastination	0.40***	0.32***	0.28***	0.07	0.05	0.61***	−0.72***	0.51***	0.50***	–						
11. Sleep procrastination	0.35***	0.29***	0.25***	0.07	0.04	0.55***	−0.60***	0.45***	0.41***	0.52***	–					
12. Salience	0.28***	0.31***	0.16***	−0.03	0.04	0.41***	−0.44***	0.42***	0.34***	0.53***	0.42***	–				
13. Reactibility	0.25***	0.37***	0.13*	0.13*	0.12*	0.29***	−0.32***	0.38***	0.32***	0.39***	0.33***	0.59***	–			
14. Monitoring	0.29***	0.36***	0.16***	−0.03	0.04	0.43***	−0.45***	0.42***	0.38***	0.53***	0.34***	0.72***	0.63***	–		
15. Age	−0.17***	−0.19***	−0.25***	−0.11*	−0.03	−0.09*	0.13**	−0.11*	−0.09	−0.26***	−0.15***	−0.18***	−0.14**	−0.19***	–	
16. Gender (1 = woman)	0.06	−0.02	0.02	0.00	−0.04	0.00	−0.03	0.04	0.08	0.01	−0.03	0.00	0.09	0.05	0.02	–

Note. Correlations between age, gender, and the daily outcome variables in the baseline week. $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table 3
Means and Standard Deviations.

	Baseline week (high personalization) M (SD)	Experimental week (low personalization) M (SD)	M_d
TikTok use			
Time spent (minutes per day)	**87.17 (80.53)**	**46.64 (60.99)**	**−40.53**
Use frequency (times per day)	**19.45 (19.02)**	**14.17 (15.65)**	**−5.28**
Entertainment experience			
Personal relevance	**5.36 (1.17)**	**3.01 (1.63)**	**−2.35**
Enjoyment	**5.33 (1.22)**	**3.49 (1.75)**	**−1.84**
Meaningfulness	**4.11 (1.32)**	**2.80 (1.51)**	**−1.31**
Time distortion	**4.37 (1.86)**	**3.05 (1.70)**	**−1.32**
Self-regulation			
Perceived control	**3.71 (2.02)**	**4.99 (1.77)**	**1.28**
Opening automaticity	**4.46 (1.85)**	**3.72 (1.92)**	**−0.74**
Scrolling automaticity	**5.25 (1.79)**	**4.08 (1.91)**	**−1.17**
Task procrastination	**3.78 (1.99)**	**2.73 (1.63)**	**−1.05**
Sleep procrastination	**3.22 (1.99)**	**2.36 (1.62)**	**−0.86**
TikTok vigilance			
Salience	**2.77 (1.60)**	**2.36 (1.46)**	**−0.41**
Reactibility	2.49 (1.59)	2.39 (1.58)	−0.10
Monitoring	**2.78 (1.56)**	**2.44 (1.47)**	**−0.34**

Note. For the variables marked in bold, the LMM analyses revealed a significant difference between the two weeks (all p's < 0.001).

impact of technology design on uncontrolled use (Flayelle et al., 2023), with algorithmic personalization, especially on TikTok, being a prime example (e.g., Zakaria, 2023). While public and scholarly attention on this topic has increased, empirical evidence on how feed personalization relates to user behavior and experiences remains limited. The current study, therefore, tested what happens to usage patterns and user experiences when the TikTok feed is no longer highly personalized. We found reductions in daily frequency and duration of TikTok use, as well as entertainment experiences and TikTok vigilance (salience and monitoring), and increased self-regulation when using a less personalized feed.

Our objective measures of TikTok use in the baseline week yielded statistics comparable to existing reports about the average TikTok user (De Marez et al., 2024; Dan, 2021; Pager, 2024). Participants used TikTok for one and a half hours per day on average and opened the app about 19 times per day. As expected, people used the app significantly less in the second week of the study, when their feeds were less personalized. The average daily TikTok screen time decreased by 40 min and app opening frequency decreased by 5 times per day. These findings indicate that the algorithmic personalization of TikTok's For You page plays a crucial role in keeping users engaged on the platform as well as making them return to the platform frequently.

To understand why people are less drawn to use or continue using TikTok when personalization is reduced, we examined users' entertainment experiences. Not surprisingly, participants reported that content on the less personalized feed matched their interests less compared to content on their highly personalized feed. They also found these generally popular videos less enjoyable to scroll through, and less meaningful, than the videos that were personally recommended to them based on their interests. These findings support the idea that users have personal preferences for specific content and that TikTok's algorithm is able to precisely tailor to these preferences (e.g., Lee et al., 2022). Further, participants reported lower time distortion when using a less personalized feed, suggesting a less engrossing entertainment experience.

In addition, we measured participants' perceptions of TikTok use self-regulation. When using the less personalized feed, they reported higher control over their TikTok use as they less often exceeded their intended usage time. Similarly, participants reported less automaticity in using TikTok, both in terms of opening the app and scrolling on their feeds. In general, it appears that people use TikTok in a more controlled manner when their feed is less personalized. This aligns with our findings that participants reported less task procrastination and sleep procrastination due to using TikTok. In a similar vein, participants perceived to have spent somewhat more time on offline social as well as solitary activities than they normally do. Although they also believed they had spent more time on other social media apps, suggesting TikTok use was merely replaced by other social media, the objective screen time data did not support this perception. It thus seems that switching to a less personalized feed makes it easier for users to regulate their TikTok use and prioritize other activities in their daily lives.

Finally, TikTok also occupied less mental space during moments when the app was not being used, as revealed by the significant reductions in cognitive salience of the app and the tendency to monitor the app. This finding is particularly intriguing, as it is widely believed that the increasingly pervasive nature of online platforms—marked by the constant potential for notifications, messages, and new relevant content—makes it challenging to mentally disengage (e.g., Klimmt et al., 2017). This cognitive salience of the online sphere has been linked to lower mental well-being, such as increased stress (Freytag et al., 2021) and mental fatigue (Van Gaeveren et al., 2024). Considering our finding of reduced TikTok salience due to lower feed personalization, it appears that algorithmic personalization imposes ongoing cognitive demands, fostering a sustained mental engagement with the platform even outside active use. This is problematic as algorithmic personalization becomes increasingly central to online platform design, and might thus put a strain on users' cognitive load and well-being.

As our study revealed several beneficial outcomes for users' self-regulation (e.g., less procrastination), some might conclude that

using this "non-personalization" setting is a suitable intervention—or disconnection strategy—for TikTok users who wish to regulate their usage. Indeed, opting out of personalization is a form of disconnection, as it entails a deliberate form of non-use of a certain feature (i.e., algorithmic feed personalization) to reduce the harms experienced by digital media use (e.g., perceived overuse; Nassen et al., 2023; Vanden Abeele et al., 2024). However, while our findings suggest that reduced personalization may be a promising approach, we identify two key concerns that limit its applicability as a broader solution in its current form.

First, although our participants were initially motivated to reduce their TikTok use, and despite the demonstrated effectiveness of reduced feed personalization, participants were unwilling to maintain their less personalized feed after the study ended. This pattern is similar to what we have seen in previous studies that have tested strategies designed to reduce the appeal of smartphones (e.g., grayscaling; Dekker & Baumgartner, 2024). Participants' reluctance may not be surprising, as entertainment is a core motive for using TikTok (e.g., Vaterlaus & Winter, 2021), and reduced personalization directly compromises this experience. In the open-ended responses from the exit survey, participants indeed seemed to reason that, after having experienced the pros and cons of the strategy, the cons (e.g., less fun) outweigh the pros (e.g., reduced usage). This suggests that disabling interest-based feed personalization is not a strategy that users are willing to adopt.

Second, by framing reduced personalization as a strategy that individuals can implement to better regulate their use, the responsibility is placed entirely on the user (e.g., Jorge et al., 2022). While providing users with options like less personalized feeds (i.e., the ability to opt out of default functionalities) can theoretically empower them to make more informed decisions about their online experiences, in practice, this approach may not be sufficient. More specifically, such options can be (too) unattractive if they compromise important benefits, as appears to be the case with reduced feed personalization. Furthermore, users may be unable to make fully informed choices that are in their best interests when constantly confronted with design elements that maximize engagement and retention (e.g., Flayelle et al., 2023). This may be particularly challenging for more vulnerable groups such as adolescents but also users who are already heavily invested in or "hooked" by the platform.

Thus, merely offering an opt-out option for interest-based feed personalization does not seem to be a comprehensive solution. Future regulations may therefore need to impose stricter constraints on platforms' use of engaging design functionalities that can undermine user autonomy. One way forward could be to explore alternative designs of social media platforms that would provide users with the best of both worlds (i.e., maximal entertainment and maximal control). While TikTok's current reduced personalization setting is not ideal as it takes away too much of the pleasure, our findings are still promising in that they suggest that even simple changes to social media platforms can promote more controlled usage without restricting access. Perhaps the solution can be found in more nuanced or refined design changes. For example, a feed could gradually present less personalized content while scrolling, naturally encouraging users to stop their scrolling session after some time. Additionally, combining personalization adjustments with changes to other engagement-driven features, such as the infinite scroll, may prove effective (Anderson & Wood, 2020; Flayelle et al., 2023). Future research should explore how such more sophisticated design changes influence the balance between users' enjoyment and control.

4.1. *Limitations and suggestions for future research*

Although this study offers valuable insights into the impact of (reduced) feed personalization on usage and user experiences, some of the underlying mechanisms require further investigation. Most importantly, the exact components of reduced feed personalization that explain the reductions in user engagement and automaticity of use remain undetermined. One plausible explanation may be found in the reduction of content-based rewards such as enjoyment. The lack of rewards may have reduced habitual TikTok use patterns, as they were no longer reinforced by positive content-based rewards (Anderson & Wood, 2020). Moreover, in addition to less rewarding content, people may encounter content they find highly unpleasant or disturbing. For example, in a qualitative study, first-time TikTok users reported that some videos on their not-yet-personalized feed made them feel uncomfortable, particularly when they contained sexualized content (Siles et al., 2024). These unpleasant video encounters may serve as stopping cues that disrupt the automated behaviors of content consumption (i.e., friction; Anderson & Wood, 2020).

At the same time, a personalized feed is not only characterized by a high number of rewarding videos but these videos are presented in an intermittently rewarding pattern. This means that more rewarding videos are alternated with less rewarding videos to maintain users' continuous curiosity for the next pleasantly surprising video (Kang & Lou, 2022; Ramsden & Talbot, 2024). This pattern contributes to habitual scrolling to the next content (e.g., Bayer et al., 2022). However, it is uncertain to what extent this variable reward system remains present in the less personalized feed, as it could, for instance, be that this pattern is mimicked by presenting both more popular videos and less popular videos. It thus remains unclear whether the reduction in automaticity of use was primarily due to a decrease in rewarding content, an increase in negative experiences, or different content delivery patterns. Further research is needed to determine the precise mechanisms that drive and disrupt habitual social media use patterns. More in general, underlying mechanisms and potential mediation effects (e.g., reduced usage leading to lower salience) warrant further research, as they may help explain how algorithmic personalization impacts user experiences.

In addition, the strong reduction in TikTok usage may partially be explained by participants' pre-existing strong intentions to reduce their app use. That is, in the entry survey, participants indicated spending too much time on the app and wishing to reduce their usage. They may, therefore, have taken the study as an opportunity to (finally) act upon their perceived TikTok overuse. It could thus be that participants responded more strongly to the reduced personalization of their feeds than they would have if they were unconcerned about their TikTok use. However, since the less personalized feed also strongly decreased enjoyment, similar effects may be expected for all users. Nevertheless, future research may investigate whether our findings generalize to people who are satisfied with their TikTok use.

Similarly, given that participants were instructed to switch to the less personalized feed after the baseline week, their responses and behavior might have been shaped by expectations about what the study was investigating. While we attempted to minimize demand characteristics by advertising the study neutrally (i.e., not as a "social media reduction experiment" but as a study testing the effects of TikTok settings on user experiences), we cannot fully eliminate the possibility of demand effects. Future research could address this concern more directly by employing alternative designs in which participants remain unaware of the experimental condition, such as lab-based studies where researchers adjust personalization settings covertly.

A related shortcoming of the present study is the absence of a control group. We opted for a within-subjects design, which enhances statistical power and feasibility of the study (i.e., lower required sample size). Still, without random assignment, alternative explanations cannot be entirely ruled out. Therefore, while our findings are robust and aligned with theoretical expectations, they are not definitively causal. Future work could employ designs that combine randomization with high ecological validity, such as randomized controlled trials, to further isolate the effects of algorithmic personalization.

Another limitation concerns the breadth of outcome variables included. The inclusion of multiple related constructs may have introduced some conceptual overlap and interpretive complexity. Future research may adopt a more targeted approach by focusing on a smaller number of distinct outcomes.

A final avenue for future research might be to examine the role of algorithmic personalization for new TikTok users. The current study focused on the effects of reduced personalization in a sample of users already accustomed to TikTok's highly personalized feed. This raises the question of how new TikTok users would experience and use the app if their feeds never became highly personalized. It is plausible that these users would find the content irrelevant and unappealing, which could discourage continued use of the app. Alternatively, they may be less critical of the content because they lack a frame of reference for personalized TikTok experiences. As a consequence, they might find a rather satisfactory balance where they do not have to miss out on using TikTok while still being able to regulate their usage effectively. Further research is needed to explore the role of (reduced) feed personalization in the experiences of new users and how feed personalization contributes to the development of habitual or uncontrolled use.

5. Conclusion

Social media platforms, and TikTok in particular, are known for their highly engaging algorithmically personalized feeds. Yet, both users and policymakers worry that these feeds might be too engaging, leading to uncontrolled use (e.g., Ramsden & Talbot, 2024). Our study is the first to experimentally investigate how users interact with a social media platform when their feed is no longer personalized based on their interests. We found reduced enjoyment and less time spent on the platform, and participants were better able to regulate their usage. Despite these beneficial outcomes, participants considered the personalization feature indispensable, with little intention to maintain a less personalized feed. These findings confirm that feed personalization plays a critical role in engaging users and suggest that reducing feed personalization may be a promising, though currently limited, approach to address uncontrolled social media use. More research is needed to disentangle underlying mechanisms and further inform policies to better protect (vulnerable) users.

CRediT authorship contribution statement

Cynthia A. Dekker: Writing – review & editing, Writing – original draft, Project administration, Methodology, Formal analysis, Conceptualization. **Susanne E. Baumgartner:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Sindy R. Sumter:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Funding

This study was funded by the Amsterdam School of Communication Research (ASCoR).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data underlying this article are available on the Open Science Framework at https://osf.io/se37j/.

References

Allyn, B., 2024 October 8. More than a dozen states sue TikTok, alleging it harms kids and is designed to addict them. NPR. https://www.npr.org/2024/10/08/g-s1-26625/states-sue-tiktok-child-safety-mental-health.
Anderson, I.A., Wood, W., 2020. Habits and the electronic herd: The psychology behind social media's successes and failures. Consum. Psychol. Rev. 4 (1), 83–99. https://doi.org/10.1002/arcp.1063.
Barr, D.J., Levy, R., Scheepers, C., Tily, H.J., 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. J. Mem. Lang. 68 (3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001.

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67 (1), 1–48. https://doi.org/10.18637/jss.v067.i01.

Baumann, F., Arora, N., Rahwan, I., Czaplicka, A., 2025. Dynamics of algorithmic content amplification on TikTok. arXiv. https://doi.org/10.48550/arXiv.2503.20231.

Bayer, J.B., Anderson, I.A., Tokunaga, R.S., 2022. Building and breaking social media habits. Curr. Opin. Psychol. 45, 101303. https://doi.org/10.1016/j.copsyc.2022.101303.

Bhandari, A., Bimo, S., 2022. Why's everyone on TikTok now? The algorithmized self and the future of self-making on social media. Soc. Media Soc. 8 (1), 1–11. https://doi.org/10.1177/20563051221086241.

Boeker, M., Urman, A., 2022. An empirical investigation of personalization factors on TikTok. Proceedings of the ACM Web Conference 2022, 2298–2309. https://doi.org/10.1145/3485447.3512102.

Bucher, T., 2017. The algorithmic imaginary: Exploring the ordinary affects of Facebook algorithms. Inf. Commun. Soc. 20 (1), 30–44. https://doi.org/10.1080/1369118X.2016.1154086.

Bucknell Bossen, C., Kottasz, R., 2020. Uses and gratifications sought by pre-adolescent and adolescent TikTok consumers. YoungConsum. 21 (4), 463–478. https://doi.org/10.1108/YC-07-2020-1186.

Chen, X., Hedman, A., Distler, V., Koenig, V., 2023. Do persuasive designs make smartphones more addictive? A mixed-methods study on Chinese university students. Comput. Hum. Behav. Rep. 10, 100299. https://doi.org/10.1016/j.chbr.2023.100299.

De Marez, L., Sevenhant, R., Denecker, F., Georges, A., Wuyts, G., Schuurman, D., 2024. Digitale trends in Vlaanderen. imec.digimeter 2023. Imec https://www.imec.be/nl/kennisontwikkeling/technotrends/digimeter/imecdigimeter-2023.

Dekker, C.A., Baumgartner, S.E., 2024. Is life brighter when your phone is not? The efficacy of a grayscale smartphone intervention addressing digital well-being. Mob. Media Commun. 12 (3), 688–708. https://doi.org/10.1177/20501579231212062.

Dong, Z., Xie, T., 2024. Why do people love short-form videos? The motivations for using Chinese TikTok (Douyin) and implications for well-being. Curr. Psychol. 43 (26), 22283–22296. https://doi.org/10.1007/s12144-024-05927-4.

Dua, T., 2021, April 13. Never-before-seen TikTok stats from leaked sales presentations show how it's trying to lure advertisers to the platform. Bus. Insid. https://www.businessinsider.com/tiktok-pitch-deck-shows-new-e-commerce-ads-2021-4.

Eg, R., Demirkol Tønnesen, Ö., Tennfjord, M.K., 2023. A scoping review of personalized user experiences on social media: The interplay between algorithms and human factors. Comput. Hum. Behav. Rep. 9, 100253. https://doi.org/10.1016/j.chbr.2022.100253.

European Commission. (2024, February 19). DSA: Commission opens formal proceedings against TikTok. https://ec.europa.eu/commission/presscorner/detail/en/IP_24_926.

Faber, A., Bee, C., Girju, M., Osel, N., Rossi, A.M., Crane, M., Lutz, R.J., Nardini, G., Jung, C.E., 2022. The paradoxes of smartphone use: Understanding the user experience in today's connected world. J. Consum. Aff. 56 (3), 1260–1280. https://doi.org/10.1111/joca.12472.

Fagan, A., 2024, April 17). TikTok leads time spent on social for teens US adults. EMARKETER https://www.emarketer.com/content/tiktok-leads-time-spent-on-social-most-us-adults.

Feldkamp, J., 2021. The rise of TikTok: The evolution of a social media platform during COVID-19. In: Hovestadt, C., Recker, J., Richter, J., Werder, K. (Eds.), Digital Responses to Covid-19. Springer, pp. 73–85. https://doi.org/10.1007/978-3-030-66611-8_6.

Flayelle, M., Brevers, D., King, D.L., Maurage, P., Perales, J.C., Billieux, J., 2023. A taxonomy of technology design features that promote potentially addictive online behaviours. Nat. Rev. Psychol. 2 (3), 136–150. https://doi.org/10.1038/s44159-023-00153-4.

Freytag, A., Knop-Huelss, K., Meier, A., Reinecke, L., Hefner, D., Klimmt, C., Vorderer, P., 2021. Permanently online—always stressed out? The effects of permanent connectedness on stress experiences. Hum. Commun. Res. 47 (2), 132–165. https://doi.org/10.1093/hcr/hqaa014.

Gardner, B., Abraham, C., Lally, P., de Bruijn, G.-J., 2012. Towards parsimony in habit measurement: Testing the convergent and predictive validity of an automaticity subscale of the Self-Report Habit Index. Int. J.Behav. Nutr. Phys. Act. 9 (1), 102. https://doi.org/10.1186/1479-5868-9-102.

Guess, A.M., Malhotra, N., Pan, J., Barberá, P., Allcott, H., Brown, T., Crespo-Tenorio, A., Dimmery, D., Freelon, D., Gentzkow, M., González-Bailón, S., Kennedy, E., Kim, Y.M., Lazer, D., Moehler, D., Nyhan, B., Rivera, C.V., Settle, J., Thomas, D.R., Tucker, J.A., 2023. How do social media feed algorithms affect attitudes and behavior in an election campaign? Science 381 (6656), 398–404. https://doi.org/10.1126/science.abp9364.

Hilbert, M., Cingel, D.P., Zhang, J., Vigil, S.L., Shawcroft, J., Xue, H., Thakur, A., Shafiq, Z., 2023. #BigTech @Minors: Social media algorithms quickly personalize minors' content, lacking equally quick protection. Soc. Sci. Res.Netw. https://doi.org/10.2139/ssrn.4674573.

Holzmann, Z. (2021, March 17). The ugly side of TikTok. Cardinal. https://wearecardinals.com/1504-thematas/the-ugly-side-of-tiktok-/.

Huang, Q., Hu, M., Zhang, N., 2022. A techno-psychological approach to understanding problematic use of short-form video applications: The role of flow. Front. Psychol. 13. https://doi.org/10.3389/fpsyg.2022.971589.

Jorge, A., Amaral, I., De Alves, A.M., 2022. "Time well spent": The ideology of temporal disconnection as a means for digital well-being. Int. J.Commun. 16, 1551–1572.

Kang, H., Lou, C., 2022. AI agency vs. human agency: Understanding human–AI interactions on TikTok and their implications for user engagement. J. Comput.-Mediat. Commun. 27 (5), zmac014. https://doi.org/10.1093/jcmc/zmac014.

Klimmt, C., Hefner, D., Reinecke, L., Rieger, D., Vorderer, P., 2017. The permanently online and permanently connected mind. Mapping the cognitive structures behind mobile internet use. In: Permanently online, permanently connected: Living and communicating in a POPC World. Routledge, pp. 18–28. https://doi.org/10.4324/9781315276472-3.

Klug, D., Qin, Y., Evans, M., & Kaufman, G. (2021). Trick and please. A mixed-method study on user assumptions about the TikTok algorithm. Proceedings of the 13th ACM Web Science Conference 2021, 84–92. https://doi.org/10.1145/3447535.3462512.

Kroese, F.M., Evers, C., Adriaanse, M.A., de Ridder, D.T., 2016. Bedtime procrastination: A self-regulation perspective on sleep insufficiency in the general population. J. Health Psychol. 21 (5), 853–862. https://doi.org/10.1177/1359105314540014.

Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B., 2017. lmerTest package: Tests in linear mixed effects models. J. Stat. Softw. 82 (13), 1–26. https://doi.org/10.18637/jss.v082.i13.

Lee, A.Y., Mieczkowski, H., Ellison, N.B., Hancock, J.T., 2022. The algorithmic crystal: Conceptualizing the self through algorithmic personalization on TikTok. In: Proceedings of the ACM on Human-Computer Interaction. https://doi.org/10.1145/3555601.

Lu, X., Lu, Z., & Liu, C. (2020). Exploring TikTok use and non-use practices and experiences in China. In G. Meiselwitz (Ed.), Social Computing and Social Media. Participation, User Experience, Consumer Experience, and Applications of Social Computing. HCII 2020 (pp. 57–70). Springer International Publishing. https://doi.org/10.1007/978-3-030-49576-3_5.

Martela, F., Ryan, R.M., 2024. Assessing autonomy, competence, and relatedness briefly: Validating single-item scales for basic psychological need satisfaction. Eur. J. Psychol. Assess. https://doi.org/10.1027/1015-5759/a000846.

Martinez, R.J., Brammer, S.E., Punyanunt-Carter, N.M., 2024. 'I can spend forever on it without getting bored': Analyzing what college students like and dislike about TikTok. Media Watch 15 (1), 113–130. https://doi.org/10.1177/09760911231211535.

Matthews, R.A., Pineault, L., Hong, Y.-H., 2022. Normalizing the use of single-item measures: Validation of the single-item compendium for organizational psychology. J. Bus. Psychol. 37 (4), 639–673. https://doi.org/10.1007/s10869-022-09813-3.

Meier, A., Reinecke, L., Meltzer, C.E., 2016. "Facebocrastination"? Predictors of using Facebook for procrastination and its effects on students' well-being. Comput. Hum. Behav. 64, 65–76. https://doi.org/10.1016/j.chb.2016.06.011.

Montag, C., Hegelich, S., 2020. Understanding detrimental aspects of social media use: Will the real culprits please stand up? Front. Sociol. 5, 599270. https://doi.org/10.3389/fsoc.2020.599270.

Nassen, L.-M., Vandebosch, H., Poels, K., Karsay, K., 2023. Opt-out, abstain, unplug. A systematic review of the voluntary digital disconnection literature. Telematics Inform. 81, 101980. https://doi.org/10.1016/j.tele.2023.101980.

Oliver, M.B., Bartsch, A., 2010. Appreciation as audience response: Exploring entertainment gratifications beyond hedonism. Hum. Commun. Res. 36 (1), 53–81. https://doi.org/10.1111/j.1468-2958.2009.01368.x.
Ramsden, E., Talbot, C.V., 2024. The role of TikTok in students' health and wellbeing. Int. J.Ment. Heal. Addict. https://doi.org/10.1007/s11469-023-01224-6.
R Core Team. (2024). R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.R-project.org/.
Reinecke, L., Klimmt, C., Meier, A., Reich, S., Hefner, D., Knop-Huelss, K., Rieger, D., Vorderer, P., 2018. Permanently online and permanently connected: Development and validation of the Online Vigilance Scale. PLoS One 13 (10), e0205384. https://doi.org/10.1371/journal.pone.0205384.
Schellewald, A., 2022. Theorizing "stories about algorithms" as a mechanism in the formation and maintenance of algorithmic imaginaries. Soc. Media Soc. 8 (1), 1–10. https://doi.org/10.1177/20563051221077025.
Schellewald, A., 2023. Understanding the popularity and affordances of TikTok through user experiences. Media Cult. Soc. 45 (8), 1568–1582. https://doi.org/10.1177/01634437221144562.
Scherr, S., Wang, K., 2021. Explaining the success of social media with gratification niches: Motivations behind daytime, nighttime, and active use of TikTok in China. Comput. Hum. Behav. 124, 106893. https://doi.org/10.1016/j.chb.2021.106893.
Sherry, J.L., 2004. Flow and media enjoyment. Commun. Theory 14 (4), 328–347. https://doi.org/10.1111/j.1468-2885.2004.tb00318.x.
Siles, I., Valerio-Alfaro, L., Meléndez-Moran, A., 2024. Learning to like TikTok and not: Algorithm awareness as process. New Media Soc. 26 (10), 5702–5718. https://doi.org/10.1177/14614448221138973.
Su, C., Zhou, H., Gong, L., Teng, B., Geng, F., Hu, Y., 2021. Viewing personalized video clips recommended by TikTok activates default mode network and ventral tegmental area. NeuroImage 237, 118136. https://doi.org/10.1016/j.neuroimage.2021.118136.
Taylor, S.H., Choi, M., 2022. An initial conceptualization of algorithm responsiveness: Comparing perceptions of algorithms across social media platforms. Soc. Media Soc. 8 (4). https://doi.org/10.1177/20563051221144322.
TikTok. (2020, June 18). How TikTok recommends videos #ForYou. https://newsroom.tiktok.com/en-us/how-tiktok-recommends-videos-for-you.
TikTok. (2023, August 28). Our compliance with the Digital Services Act. https://newsroom.tiktok.com/en-eu/compliance-digital-services-act-eu.
Van Gaeveren, K., Murphy, S.L., de Segovia Vicente, D., Vanden Abeele, M.M.P., 2024. Connected yet cognitively drained? A mixed-methods study examining whether online vigilance and availability pressure promote mental fatigue. Commun. Res. https://doi.org/10.1177/00936502241298494.
Vanden Abeele, M.M.P., 2021. Digital wellbeing as a dynamic construct. Commun. Theory 31 (4), 932–955. https://doi.org/10.1093/ct/qtaa024.
Vanden Abeele, M.M.P., Nguyen, M.H., 2024. Digital media as ambiguous goods: Examining the digital well-being experiences and disconnection practices of Belgian adults. Eur. J. Commun. 39 (2), 122–144. https://doi.org/10.1177/02673231231201487.
Vanden Abeele, M.M.P., Vandebosch, H., Koster, E.H.W., De Leyn, T., Van Gaeveren, K., de Segovia Vicente, D., Van Bruysseel, S., van Timmeren, T., De Marez, L., Poels, K., DeSmet, A., De Wever, B., Verbruggen, M., Baillien, E., 2024. Why, how, when, and for whom does digital disconnection work? A process-based framework of digital disconnection. Commun. Theory 1–15. https://doi.org/10.1093/ct/qtad016.
Vaterlaus, J.M., Winter, M., 2021. TikTok: An exploratory study of young adults' uses and gratifications. Soc. Sci. J. 1–20. https://doi.org/10.1080/03623319.2021.1969882.
Vombatkere, K., Mousavi, S., Zannettou, S., Roesner, F., Gummadi, K.P., 2024. TikTok and the art of personalization: investigating exploration and exploitation on social media feeds. Proceed. ACM on Web Conference 2024, 3789–3797. https://doi.org/10.1145/3589334.3645600.
Vorderer, P., Klimmt, C., Ritterfeld, U., 2004. Enjoyment: At the heart of media entertainment. Commun. Theory 14 (4), 388–408. https://doi.org/10.1111/j.1468-2885.2004.tb00321.x.
Wang, K., Scherr, S., 2022. Dance the night away: how automatic TikTok use creates pre-sleep cognitive arousal and daytime fatigue. Mob. Media Commun. 10 (2), 316–336. https://doi.org/10.1177/20501579211056116.
Wolfers, L.N., Baumgartner, S.E., Zhang, S.E., Yang, H., 2023. Short but still valid: Validating single-item measures for key media psychology constructs for experience sampling research. In: Proceedings of the 13th Conference of the Media Psychology Division (DGPs). Melusina Press, pp. 77–78. https://doi.org/10.26298/1981-5555.
Xie, J., Xu, X., Zhang, Y., Tan, Y., Wu, D., Shi, M., Huang, H., 2023. The effect of short-form video addiction on undergraduates' academic procrastination: A moderated mediation model. Front. Psychol. 14, 1298361. https://doi.org/10.3389/fpsyg.2023.1298361.
Yang, H., Zhang, S., Diao, Z., Sun, D., 2023. What motivates users to continue using current short video applications? A dual-path examination of flow experience and cognitive lock-in. Telematics Inform. 85, 102050. https://doi.org/10.1016/j.tele.2023.102050.
Ytre-Arne, B., Syvertsen, T., Moe, H., Karlsen, F., 2020. Temporal ambivalences in smartphone use: Conflicting flows, conflicting responsibilities. New Media Soc. 22 (9), 1715–1732. https://doi.org/10.1177/1461444820913561.
Zakaria, F. (2023, April 14). TikTok is dangerously addictive. We should regulate it now. Washington Post. https://www.washingtonpost.com/opinions/2023/04/14/tiktok-teens-mental-health-regulation/.
Zhao, H., Wagner, C., 2022. How TikTok leads users to flow experience: investigating the effects of technology affordances with user experience level and video length as moderators. Internet Res. 33 (2), 820–849. https://doi.org/10.1108/INTR-08-2021-0595.
Zhou, Z. (2021). Analysis on the "Douyin (TikTok) mania" phenomenon based on recommendation algorithms. 2020 International Conference on New Energy Technology and Industrial Development (NETID 2020), 235. https://doi.org/10.1051/e3sconf/202123503029.

Body Image 41 (2022) 292–297

Contents lists available at ScienceDirect

Body Image

journal homepage: www.journals.elsevier.com/body-image

The dangers of the rabbit hole: Reflections on social media as a portal into a distorted world of edited bodies and eating disorder risk and the role of algorithms

Jennifer A. Harriger[a,*], Joshua A. Evans[a], J. Kevin Thompson[b], Tracy L. Tylka[c]

[a] Social Science Division, Pepperdine University, Malibu, CA, USA
[b] Department of Psychology, University of South Florida, Tampa, FL, USA
[c] Department of Psychology, Ohio State University, Columbus, OH, USA

ARTICLE INFO

Article history:
Received 22 March 2022
Accepted 23 March 2022
Available online 1 April 2022

Keywords:
Social media
Algorithms
Corporate social responsibility
Social media influencers
Social action

ABSTRACT

The relationship between social media usage and body image has been well-established in the literature; however, social media companies' use of algorithms may intensify this association, as algorithms provide viewers with personalized content that is often more extreme, less monitored, and designed to keep users engaged for longer periods of time. This article details the recent media coverage of algorithms, revelations by former social media employees regarding the problematic usage of algorithms, and revelations that social media companies are aware of the harm posed by their implementation of algorithms, particularly for young, vulnerable users. We provide recommendations for influencers, educators, researchers, clinicians, parents, and users, and conclude that it is ultimately the responsibility of the social media corporations to protect and enhance the well-being of their users.

Contents

Engagement algorithm

1. Social media and body image

Much like teenage Alice who entered a distorted wor… falling down a rabbit hole in *Alice in Wonderland* and view… body in a skewed manner in comparison to others, a parallel…

…lescents and young …bbit hole users into …s that may prompt …ed concerns, eating …z-Rodriguez, Verde- … Several reviews have unvaryingly concluded that social media use is linked to higher body dissatisfaction (Fardouly & Vartanian, 2016; Saiphoo & Vahedi, 2019). While earlier research focused on the overall time spent on social

* Corresponding author.
E-mail address: jennifer.harriger@pepperdine.edu (J.A. Harriger).

https://doi.org/10.1016/j.bodyim.2022.03.007

Body Image 41 (2022) 292–297

Contents lists available at ScienceDirect

Body Image

journal homepage: www.journals.elsevier.com/body-image

The dangers of the rabbit hole: Reflections on social media as a portal into a distorted world of edited bodies and eating disorder risk and the role of algorithms

Jennifer A. Harriger[a,*], Joshua A. Evans[a], J. Kevin Thompson[b], Tracy L. Tylka[c]

[a] Social Science Division, Pepperdine University, Malibu, CA, USA
[b] Department of Psychology, University of South Florida, Tampa, FL, USA
[c] Department of Psychology, Ohio State University, Columbus, OH, USA

ARTICLE INFO

Article history:
Received 22 March 2022
Accepted 23 March 2022
Available online 1 April 2022

Keywords:
Social media
Algorithms
Corporate social responsibility
Social media influencers
Social action

ABSTRACT

The relationship between social media usage and body image has been well-established in the literature; however, social media companies' use of algorithms may intensify this association, as algorithms provide viewers with personalized content that is often more extreme, less monitored, and designed to keep users engaged for longer periods of time. This article details the recent media coverage of algorithms, revelations by former social media employees regarding the problematic usage of algorithms, and revelations that social media companies are aware of the harm posed by their implementation of algorithms, particularly for young, vulnerable users. We provide recommendations for influencers, educators, researchers, clinicians, parents, and users, and conclude that it is ultimately the responsibility of the social media corporations to protect and enhance the well-being of their users.

Contents

1. Social media and body image

Much like teenage Alice who entered a distorted world after falling down a rabbit hole in *Alice in Wonderland* and viewed her body in a skewed manner in comparison to others, a parallel can be argued in the journey many contemporary adolescents and young adults take into social media platforms that rabbit hole users into emotionally extreme content and edited bodies that may prompt mental health risks such as appearance-related concerns, eating disorders, and body dysmorphia (Padin, Gonzalez-Rodriguez, Verde-Diego & Vazquez-Perez, 2021). Several reviews have unvaryingly concluded that social media use is linked to higher body dissatisfaction (Fardouly & Vartanian, 2016; Saiphoo & Vahedi, 2019). While earlier research focused on the overall time spent on social

* Corresponding author.
E-mail address: jennifer.harriger@pepperdine.edu (J.A. Harriger).

https://doi.org/10.1016/j.bodyim.2022.03.007

media, more recent research has reported that photo or appearance-based activities (e.g., posting photos or commenting on others' photos) have been found to have a deleterious effect on body image (for reviews, see Fardouly & Vartanian, 2016; Holland & Tiggemann, 2016; Saiphoo & Vahedi, 2019). Additionally, researchers' findings regarding the negative impacts of taking and editing selfies (Lonergan et al., 2020; Mills, Musto, Williams & Tiggemann, 2018; Tiggemann et al., 2020; Vandenbosch et al., 2022; Wang, Xie, Fardouly, Vartanian & Lei, 2021; Wick & Keel, 2020) have also led to increased concern regarding the effects of social media usage in adolescents and young adults.

The majority of studies published before 2017 focused primarily on Facebook, while newer research has included a variety of other platforms. Platforms such as Instagram and Snapchat, which are more photo-based than platforms such as Facebook and Twitter, are more likely to be linked to body dissatisfaction (Karsay, Trekels, Eggermont & Vandenbosch, 2021; Willoch, O'Shea, Ho, Byrne & Wade, 2020). Finally, platforms such as TikTok (which is primarily video-based) are only beginning to be examined by researchers (Vandenbosch et al., 2022). TikTok appeals to a younger demographic; approximately 25% of users in the US are between the ages of 10–19 years of age (Statista, 2022), and this age group is particularly vulnerable to the effects of social media (Vandenbosch et al., 2022) and may be less likely to recognize that the images they interact with are often digitally manipulated.

2. Use of editing within social media

Social media content is frequently filtered and edited by computer software programs which contribute to the unrealistic standards of beauty portrayed on social media (Brown, 2020; Grindell, 2021). These software programs can be used to modify photos and videos in ways that users often are unable to detect as edited, and many users report surprise that editing physiques on videos was even possible (Liotta, 2021). Many social media influencers (i.e., a user high in social standing who has the power to affect their followers' beliefs and purchasing decisions) showcase highly edited bodies they claim they achieved through diet, exercise, or products they are paid to promote, and exposure to unrealistic idealized images is linked to increased risk for disordered eating and body dissatisfaction through mechanisms such as self-objectification and appearance comparisons (for a review of the literature, see Vandenbosch et al., 2022). In response to unrealistic images, online communities dedicated to the promotion of body positivity have grown. These platforms provide users with strategies to resist the pursuit of unrealistic appearance ideals and to appreciate both the appearance and functionality of their own and others' bodies (Cohen et al., 2019; Lazuka et al., 2020); yet little is known regarding the percentage of users that follow these types of accounts compared to, or even in addition to, accounts with idealized imagery.

3. The role of algorithms

While social media offers users the ability to play an active role in determining the type of content they are exposed to, such as the choice of which accounts to follow (Perloff, 2014), users are also often unknowingly affected by social media algorithms which bias the content they are shown (Wall Street Journal, 2021). This situation means that even users who may make efforts to avoid certain types of content or to follow more positive accounts are not always protected. The algorithm generated content for each user tends to be personalized, often including high levels of advertising where it is often unclear whether the posts contain user-contributed or commercial content (Perloff, 2014). Additionally, some social media platforms, such as Tik Tok, tend to rabbit hole users into less monitored and more extreme content quickly, as the algorithm learns what users are willing to view for longer, in an attempt to increase the time they spend on the platform (Wall Street Journal, 2021).

Social media companies are aware of the harm being caused by their current content approaches. Recent whistleblower revelations demonstrated that social media giant Meta (formerly known as Facebook) documented the company's knowledge of the negative mental health effects of their products, such as Facebook and Instagram, on young people (Wells et al., 2021). Former Facebook executive Frances Haugen made public internal Facebook documents detailing internal research that highlighted that the company was aware of the negative impact its social media products were having on its teenage users' mental health and body image (Mac & Kang, 2021; Wells, et al., 2021). Confidential Facebook research leaked to *The Wall Street Journal* detailed a connection between Instagram use and negative body image in one out of three of the teenage girls they studied (Wells, et al., 2021). Haugen's documents have been shared with Congress, state attorney generals, and media sources (Mac & Kang, 2021).

The Wall Street Journal also conducted its own experiment in which they created over 100 new bot run Tik Tok accounts, and they found these accounts quickly tended to rabbit hole users into more niche intense content which was less moderated by Tik Tok staff, and which featured emotionally triggering material, such as discussions and behaviors which could negatively impact mental health (e.g., videos about self-harm behaviors; Wall Street Journal, 2021). Tik Tok sends users recommendations in their "For You" tab that mixes new content a user might enjoy with other videos that person might not usually independently seek out or view (Hu, 2021), and many users may stumble upon less monitored content via the use of this tab. Officially, TikTok stated preferences such as preferred language, user's geographic location, time spent on certain videos, comments, likes, and the types of accounts a user followed were all weighted to decide what content is added to a user's "For You" suggestions (TikTok, 2020). However, in their research, *The Wall Street Journal* found the most important metric that appeared to mitigate what was next added into the "For You" tab was time spent watching particular videos, such as whether the bot paused, reviewed, left, or clicked on a type of video (Wall Street Journal Staff, 2021). Additionally, the results of the Wall Street Journal's experiment were confirmed by a whistleblower who informed *The New York Times* that the primary factor employed by the TikTok algorithm was time spent on specific types of videos (Smith, 2021).

This same Tik Tok whistleblower within the organization leaked internal corporate documents to *The New York Times* demonstrating how the company's algorithm aggressively manipulates and restricts what viewers see, with the goal being to retain viewers on the platform for longer through emotionally triggering content (Smith, 2021). That source claimed their concerns over content which promoted self-harm behaviors prompted the release of the documents (Smith, 2021). In response to criticisms, TikTok instituted content moderators to review and remove those videos which violate corporate rules and policies; however, videos in languages other than English are less monitored, moderating videos has become increasingly automated, and some triggering videos slip through TikTok's filters. Additionally, as users spend additional time on the platform, they are more likely to be exposed to non-monitored content (Hu, 2021; Kaiserman, 2021).

It has also been reported that social media use can be addictive for some adolescents (Griffiths & Kuss, 2017), and emotionally triggering content can be difficult to escape on social media, particularly when algorithms are specifically built to keep users engaged with the content that may end up being most damaging to them (Smith, 2021). Some teenagers report it is challenging to quit or reduce their social media use given that a significant amount of adolescent socialization is conducted over social media platforms, and they feared missing out on important social opportunities if they did not

frequently check their social media accounts (Franchina, Abeele, Kooij, Coco & Marez, 2018). Also, perhaps in response to limited opportunities for in-person socialization related to the COVID-19 pandemic, adolescents currently report spending more time online than in the past (Nagata et al., 2021). Furthermore, it is unrealistic to assume that children are not using social media. Even when companies implement "firm" age limits, there is nothing that prevents a younger user from creating accounts with false ages, and in fact, a recent article in the *New York Times* speculates that a third of TikTok users may be under the age of 14 (Zhong & Frenkel, 2020).

4. Recommended actions

Given the widespread use of social media, the documented negative impact of appearance-focused social media on users, and that global social media corporations are aware of this negative impact on users, we implore those in positions of power to reconsider how they implement algorithms and take action to minimize risk. In November of 2021, the Academy for Eating Disorders (AED), the International Governing Body for the research, treatment, and prevention of Eating Disorders, published a position statement asking social media companies to increase transparency around the use of the algorithms and to make community guidelines regarding appropriate content as well as paths to report content more accessible for users. They also recommended that social media companies allot resources to identify or remove accounts that promote eating disordered or weight-biased content. They urged social media companies to partner with organizations, such as AED, who can provide guidance and expert input. Finally, they recommended that researchers pre-register their upcoming studies regarding the impact of social media use on disordered eating behaviors and to share their results more publicly (Academy for Eating Disorders, 2021). We echo these recommendations, as social media corporations have a responsibility to protect users and to minimize risk, which we elaborate on below.

We offer additional recommended actions for social media corporations (including those with the power to regulate them) as well as for influencers, educators, researchers, clinicians, parents, and users. It is important to take a systems-level, macro-to-micro approach that places more responsibility on the social media corporations that create and implement the algorithms, than simply offering strategies for users to offset the influence of the algorithms that are not transparent in the first place.

4.1. For social media corporations

After Frances Haugen urged Congress to regulate social media's use of algorithms when she testified before the U.S. Senate, legislators have shown increasing interest in passing a bill that would hold social media corporations more accountable for the content they amplify using algorithms (Zakrzewski, 2021). More specifically, regulating the algorithm centers on Congress reforming Section 230, a part of the U.S. Communications Decency Act of 1996 that suggests that social media networks are not personally responsible for the content posted on their platforms. The key to successfully reforming Section 230 is ensuring the change does not include any constraint to speech but instead removes the protections social media companies enjoy around serving unrequested, emotionally laden, and at times illegal content to their users. This change may incentivize social media corporations to implement a non-algorithmic feed, such as a system where users have more control over the content they see. Even if algorithms are not ultimately removed, at minimum social media corporations should be required to (a) release details of their algorithms and core functions to be vetted by researchers, (b) disclose in users' news feeds why content was chosen, and (c) limit micro-targeting advertising messages, which use consumer data (e.g., gender, age) to choose which products to advertise to whom (e.g., diet products) (Fox, 2020).

We further advocate for corporate social responsibility within social media, which would include a social media corporation's actions that contribute to its users' well-being beyond its economic interests and legal commitments (Carroll, 2015). At minimum, social media corporations should minimize any harm that results from users engaging on their platforms, which may feature developing and implementing voluntary codes of conduct (Mayrig, 2006). In the case of appearance-focused social media, this could include working with body image and eating disorder experts to identify potentially triggering content, voluntarily providing users with transparency in how algorithms are used, and not directing users to this emotionally charged content.

We argue that social media corporations should go beyond the minimum (mitigating harm) and strive toward contributing to the well-being of its users (Porter & Kramer, 2006). For example, social media corporations could partner with body image and eating disorder experts to develop initiatives to increase user awareness of weight stigma and weight self-stigma, the harms linked to these stigmas, and how to reduce them. Likewise, this partnership could include generating ideas for how social media corporations can use their platforms to promote positive body image among users. For example, social media content identified as "body positive," which encourages body acceptance and discourages the normalization and pursuit of striving towards unrealistic appearance ideals, can be promoted on the platforms. (Of note, we strongly underscore the need to work with body image and eating disorder experts on these initiatives, as certain initiatives that the lay public believes could promote body-related well-being may have the opposite effect. For example, not all body positive content is positive or even neutral for users, as some content may result in body-related distress in viewers, see Rodgers et al., 2022).

4.2. For influencers

Social media influencers who promote body-related messages (i.e., those that center on the body, such as appearance-related and functionality-related messages) have the power to influence their followers' body image and eating behaviors. Many social media influencers are employed by companies to advertise the company's products, and they use the trust built through their social relationships and connections with their followers to sell these products. Therefore, social media influencers also have social responsibility, as they have the power to affect their followers' beliefs, self-care (or, conversely, self-destructive) behaviors, and purchasing decisions.

Given this responsibility, social media influencers should not represent companies that sell weight-control products (or any other self-destructive methods such as rigid dieting and excessive exercise) but instead select marketing messages that are body positive (e.g., promote body acceptance, body respect, and love for the body) and speak out against body shame. Like corporations, we encourage social media influencers to collaborate with body image and eating disorder experts for guidance on how to dialog with their followers to promote rather than harm their body-related well-being. For example, posting idealized appearance-focused images of themselves is likely to promote higher body dissatisfaction and greater negative mood in their followers (Lowe-Calverley & Grieve, 2020).

Importantly, many social media influencers have spoken out about the filters, body positioning, and other photo-based techniques that are used to alter appearance to attempt to increase their followers' media literacy and reduce any negative impact from viewing idealized appearance imagery. Indeed, while viewing idealized selfies has been shown to increase female followers' dissatisfaction with their face, viewing "no-makeup selfies" reduced this negative impact, suggesting that no-makeup selfies can help

educate women on the idealized nature of social media images and preserve their body image. In contrast, researchers have found that self-disclaimers (i.e., influencers' captions about the inauthenticity of their appearance) do not appear to be effective at protecting users from the harmful effects of unrealistic appearance ideals on social media (Livingston, Holland, & Fardouly, 2020). Again, we underscore the importance of influencers consulting with body image experts to identify effective and ineffective ways to offset the potential negative impact of their posts on users' body image and mood, because not all methods achieve this goal. Social media influencers could also inform their followers about algorithms and empower their followers to engage in strategies that allow algorithms to work for them (in terms of cultivating positive body image) rather than against them (leading them to appearance-focused messages). For example, influencers could encourage their followers to befriend/follow reputable body positive individuals and engage with the content on the posts.

4.3. *For researchers, educators, and clinicians*

Researchers, educators, and clinicians need to investigate effective ways to prevent the detrimental impacts of social media on users' body image and well-being and treat those with body image issues linked to social media use. Social media literacy programs could help to achieve these goals. These programs could be designed, studied, and implemented by educators, researchers, and clinicians in their work with children, adolescents, and young adults, as there is "modest, preliminary support" for the effectiveness of these programs in adolescent girls and young adult women, but not boys or young men (Paxton et al., 2022; Tamplin, McLean, & Paxton, 2018). More specifically, these programs promote critical thinking about social media via being empowered with the knowledge and skills to analyze, evaluate, produce, and participate in social media (Tamplin et al., 2018). In relation to appearance-focused social network platforms such as Instagram, social media literacy could include (a) understanding motivations for, and techniques of, commercial images and advertising as well as (b) understanding motivations for influencer and friend postings, the use of filters, and the selection and modification (editing) of images users post (McLean, Wertheim, Masters & Paxton, 2017).

One potential social media literacy intervention is *SoMe*: a 4-session program delivered in the classroom which contains activities that (a) engage participants in critiquing social media advertising and creating positive social media, (b) improve their media literacy skills in relation to realism and representation, (c) reduce the persuasive impact of social media content, and (d) lower the amount of time curating their online profile (Gordon et al., 2020). Pilot data are encouraging: participants in the intervention group demonstrated improvements in body image, dietary restraint, and media literacy relative to the control group from pre- to post-test. A more rigorous evaluation (i.e., randomized controlled trial) of whether the intervention can mitigate the negative impacts of social media engagement on early adolescent girls' and boys' body dissatisfaction, dietary restraint, depression, strategies to increase muscle, and self-esteem is underway (for additional details of the program, see Gordon et al., 2020).

It is imperative to integrate content on algorithms to social media literacy programs like SoMe. While covering algorithms and how they are used and determined is necessary, it is not sufficient. When reviewing algorithms and other content, social media literacy programs need to emphasize advocacy, empowerment, and active engagement rather than passive learning (Tiggemann, 2022). Users must be equipped with tools to combat the effects of algorithms. For example, users can learn how to exert the control they have over what information algorithms direct them toward. Such strategies could include practicing how to report and/or hide posts and overloading their feed with posts that bring them joy and are otherwise conducive to their well-being. Social media literacy programs must also stay on top of any new changes to algorithms that social media platforms implement, as the suggested strategies may then need to change. Clinicians who treat individuals with body image concerns could assess their clients' social media use and incorporate activities from social media literacy programs (with documented effectiveness) into their individual and group work. Researchers, educators, and clinicians could further communicate effective strategies to parents of young social media users.

4.4. *For parents*

With regard to social media use, parents can support the body image of their children in several ways. First, parents can examine how they themselves engage in social media. Parents can model healthy relationships with social media to their younger children (e.g., limiting the time they spend on social media or engaging with content on the Internet, following body positive content, not following content that idealizes sociocultural appearance ideals). Research indicates that parents spend significant amounts of time on their own smartphones and are often distracted by their electronic devices when interacting with their children (Auxier et al., 2020). Parental modeling of healthier relationships with electronic devices may be an important first step. Parents can also begin discussing the unrealistic images present on social media, the tools that are often used to edit the images, and the algorithms that are designed to keep users engaged for longer periods of time well before their children create their own social media accounts. This information can better equip children to navigate the challenges associated with social media use.

We also recommend that parents create social media contracts with their children that outline the rules and regulations associated with social media usage as well as consequences for lack of compliance. Family media use plans that describe the expectations related to social media use based on the age or developmental level of their children (no electronic devices at bedtime or at mealtimes, no social media an hour before bed, media free zones in the home; Derenne & Beresin, 2018) are also recommended. Finally, we also suggest that parents utilize active mediation strategies, such as discussing the risks posed by social media, rather than merely limiting social media use, which has proven less effective (Daneels & Vanwynsberghe, 2017).

4.5. *For users*

Our discussion of each system above provides guidance for individual users. Again, we reinforce that system-level change needs to occur so that individual users can effectively do their part in preserving their own body image and well-being. Social media corporations need to be transparent about how content is delivered if algorithms continue to be used, and they need to provide users with clear ways to easily opt out of content that they do not wish to see. Social media literacy programs need to have participants practice strategies that they can use to remove content that harms their body image and well-being, as well as select content that improves their body image and well-being. Only when users are armed with this knowledge and practice can they make the choice to implement strategies to preserve their body-related well-being when using social media.

5. Conclusion

Social media use is linked to higher body dissatisfaction (Fardouly & Vartanian, 2016; Saiphoo & Vahedi, 2019), and the use of algorithms serves to exacerbate this relationship by providing

viewers with personalized content that is often less monitored, more extreme, and designed to keep the user engaged for longer periods of time (Wall Street Journal, 2021). While we offer recommendations for influencers, educators, researchers, clinicians, parents, and users, we also argue that it is ultimately the responsibility of the social media corporations that create and implement the algorithms to protect their users from harm and to ultimately strive to enhance the well-being of their users.

Conflict of interest statement

We have no conflicts of interest to report.

References

Academy for Eating Disorders. (2021). Urgent responsibility to reduce harms posed by social media on risk for eating disorders: An open letter to Facebook, Instagram, TikTok, and other global social media corporations. ⟨https://www.aedweb.org/articles/urgent-responsibility-to-reduce-harms-posed-by-social-media-on-risk-for-eating-disorders⟩.

Auxier, B., Anderson, M., Perrin, A., & Turner, E. (2020, 28). Parenting Children in the Age of Screens. Pew Research Center. https://www.pewresearch.org/internet/2020/07/28/parenting-children-in-the-age-of-screens/.

Brown, D. (2020, July 29). Facetune maker launches new app that brings facial retouching to video. USA Today. ⟨https://www.usatoday.com/story/tech/2020/07/29/new-facetune-app-lets-you-retouch-your-face-videos/5534080002/⟩.

Carroll, A. B. (2016). Carroll's pyramid of CSR: Taking another look. International Journal of Corporate Social Responsibility, 1(1), 3. https://doi.org/10.1186/s40991-016-0004-6

Cohen, R., Irwin, L., Newton-John, T., & Slater, A. (2019). # bodypositivity: A content analysis of body positive accounts on Instagram. Body Image, 29, 47–57. https://doi.org/10.1016/j.bodyim.2019.02.007

Daneels, R., & Vanwynsberghe, H. (2017). Mediating social media use: Connecting parents' mediation strategies and social media literacy. article 5 Cyberpsychology: Journal of Psychosocial Research on Cyberspace, 11(3). https://doi.org/10.5817/CP2017-3-5

Derenne, J., & Beresin, E. (2018). Body image, media, and eating disorders: A 30-year update. Academic Psychiatry, 42, 129–134. https://doi.org/10.1007/s40596-017-0832-z

Fardouly, J., & Vartanian, L. R. (2016). Social media and body image concerns: Current research and future directions. Current Opinion in Psychology, 9, 1–5. https://doi.org/10.1016/j.copsyc.2015.09.005

Fox, C. (2020, November). Social media: How might it be regulated? BBC News. ⟨https://www.bbc.com/news/technology-54901083⟩.

Franchina, V., Abeele, M. V., Rooij, A. J., Coco, G. L., & Marez, L. D. (2018). Fear of missing out as a predictor of problematic social media use and phubbing behavior among Flemish adolescents. International Journal of Environmental Research and Public Health, 15(10), 2319. https://doi.org/10.3390/ijerph15102319

Gordon, C. S., Rodgers, R. F., Slater, A. E., McLean, S. A., Jarman, H. K., & Paxton, S. J. (2020). A cluster randomized controlled trial of the SoMe social media literacy body image and wellbeing program for adolescent boys and girls: Study protocol. Body Image, 33, 27–37. https://doi.org/10.1016/j.bodyim.2020.02.003

Griffiths, M. D., & Kuss, D. (2017). Adolescent social media addiction (revisited). Education and Health, 35(3), 49–52. ⟨https://www.researchgate.net/publication/320146297_Adolescent_social_media_addiction_revisited⟩.

Griodell, S. (2021, October 28). A woman's side by side TikToks show how easily videos can be edited to create unrealistic beauty standards. Insider. ⟨https://www.insider.com/woman-social-media-vs-reality-tiktoks-videos-altered-2021-10⟩.

Holland, G., & Tiggemann, M. (2016). A systematic review of the impact of the use of social networking sites on body image and disordered eating outcomes. Body Image, 17, 100–110. https://doi.org/10.1016/j.bodyim.2016.02.008

Hu, C. (2021, December 7). A look inside TikTok's seemingly all-knowing algorithm. Popular Science. ⟨https://www.popsci.com/technology/tiktok-algorithm/⟩.

Karsay, K., Trekels, J., Eggermont, S., & Vandenbosch, L. (2021). "I (don't) respect my body": Investigating the role of mass media use and self-objectification on adolescents' positive body image in a cross-national study. Mass Communication and Society, 24, 57–84. https://doi.org/10.1080/15205436.2020.1827432

Kastrenakes, J. (2021, July 9). TikTok will start automating video removals for nudity and more in the US. The Verge. ⟨https://www.theverge.com/2021/7/9/22569854/tiktok-automated-moderation-video-removal-nudity-minor-safety⟩.

Lisitsa, A. (2021, May 19). People are editing videos of their bodies to appear thinner, and this woman's viral example is starting a conversation. Buzzfeed. ⟨https://www.buzzfeed.com/alexalisitsa/body-video-editing⟩.

Lazuka, R. F., Wick, M. R., Keel, P. K., & Harriger, J. A. (2020). Are we there yet? Progress in depicting diverse images of beauty in Instagram's body positivity movement. Body Image, 34, 85–93. https://doi.org/10.1016/j.bodyim.2020.05.001

Livingston, J., Holland, E., & Fardouly, J. (2019). Exposing digital posing: The effect of social media self-disclaimer captions on women's body dissatisfaction, mood, and impressions of the user. Body Image, 32, 150–154. https://doi.org/10.1016/j.bodyim.2019.12.006

Lonergan, A. R., Bussey, K., Fardouly, J., Griffiths, S., Murray, S. B., Hay, P., Mond, J., Trompeter, N., & Mitchison, D. (2020). Protect me from my selfie: Examining the association between photo-based social media behaviors and self-reported eating disorders in adolescence. International Journal of Eating Disorders, 53(5), 755–766. https://doi.org/10.1002/eat.23256

Lowe-Calverley, E., & Grieve, R. (2020). Do the metrics matter? An experimental investigation of Instagram influencer effects on mood and body dissatisfaction. Body Image, 36, 1–4. https://doi.org/10.1016/j.bodyim.2020.10.003

Mac, R., & Kang, C. (2021, October 27). Whistle-blower says facebook 'chooses profits over safety'. The New York Times. ⟨https://www.nytimes.com/2021/10/03/technology/whistle-blower-facebook-frances-haugen.html⟩.

McLean, S. A., Wertheim, E. H., Masters, J., & Paxton, S. J. (2017). A pilot evaluation of a social media literacy intervention to reduce risk factors for eating disorders. International Journal of Eating Disorders, 50, 847–851. https://doi.org/10.1002/eat.22708

Mills, J. S., Musto, S., Williams, L., & Tiggemann, T. (2018). Selfie harm: Effects on mood and body image in young women. Body Image, 27, 86–92. https://doi.org/10.1016/j.bodyim.2018.08.007

Nagata, J. M., Cortez, C. A., & Cattle, C. J. (2021). Screen time use among us adolescents during the COVID-19 pandemic: Findings from the Adolescent Brain Cognitive Development (ABCD) study. JAMA Pediatrics, 176(1), 94–96. https://doi.org/10.1001/jamapediatrics.2021.4334

Padín, P. F., González-Rodríguez, R., Verde-Diego, C., & Vazquez-Perez, R. (2021). Social media and eating disorder psychopathology: A systematic review. Cyberpsychology: Journal of Psychosocial Research on Cyberspace, 15(3), 6. https://doi.org/10.5817/CP2021-3-6

Paxton, S. J., McLean, S. A., & Rodgers, R. F. (2022). "My critical filter buffers your app filter": Social media literacy as a protective factor for body image. Body Image, 40, 158–164. https://doi.org/10.1016/j.bodyim.2021.12.009

Perloff, R. M. (2014). Social Media Effects on Young Women's Body Image Concerns: Theoretical Perspectives and an Agenda for Research. Sex Roles, 71, 363–377. https://doi.org/10.1007/s11199-014-0384-6

Porter, M., & Kramer, M. (2006). Strategy and society: The link between corporate social responsibility and competitive advantage. Harvard Business Review, 84, 78–92. Retrieved from https://hbr.org.

Rodgers, R. F., Wertheim, E. H., Paxton, S. J., Tylka, T. L., & Harriger, J. A. (2022). #Bopo: Enhancing Body Image Through Body Positive Social Media- Evidence to Date and Research Directions. Body Image, Advance online publication.

Saiphoo, A. N., & Vahedi, Z. (2019). A meta-analytic review of the relationship between social media use and body image disturbance. Computers in Human Behavior, 101, 259–275. https://doi.org/10.1016/j.chb.2019.07.028

Smith, B. (2021, December 5). How TikTok reads your mind. The New York Times. ⟨https://www.nytimes.com/2021/12/05/business/media/tiktok-algorithm.html?smid=fb-nytimes&smtyp=cur&fbclid=IwAR3...⟩.

Statista Research Department (2022). Distribution of TikTok users in the United States, as of September 21, 2021. ⟨https://www.statista.com/statistics/1095186/tiktok-us-users-age/⟩.

Tamplin, N. C., McLean, S., & Paxton, S. J. (2018). Social media literacy protects against the negative impact of exposure to appearance ideal social media images in young adult women but not men. Body Image, 26, 29–37. https://doi.org/10.1016/j.bodyim.2018.05.003

Tiggemann, M., Anderberg, I., & Brown, Z. (2020). Uploading your best self: Selfie editing and body dissatisfaction. Body Image, 33, 175–182. https://doi.org/10.1016/j.bodyim.2020.03.002

TikTok. (2020, June 18). How TikTok recommends videos #ForYou. ⟨https://newsroom.tiktok.com/en-us/how-tiktok-recommends-videos-for-you⟩.

Tiggemann, Marika (2022). #Bopo: Enhancing Body Image Through Body Positive Social Media- Evidence to Date and Research Directions. Body Image, 41, 172–180. https://doi.org/10.1016/j.bodyim.2022.02.012

Vandenbosch, L., Fardouly, J., & Tiggemann, M. (2022). Social media and body image: Recent trends and future directions. Current Opinion in Psychology, 45, Article 101289. https://doi.org/10.1016/j.copsyc.2021.12.002

Wall Street Journal. (2021, July 21). Investigation: How TikTok's algorithm figures out your deepest desires. The Wall Street Journal. ⟨https://www.wsj.com/video/series/inside-tiktoks-highly-secretive-algorithm/investigation-how-tiktok-algorithm-figures-out-your-deepest-desires/6C0C2040-FF25-4827-8528-2BD6612D3796?mod=djp_lead_pos1&mod=article_inline⟩.

Wall Street Journal Staff. (2021, July 21). Inside TikTok's algorithm: A WSJ Investigation. The Wall Street Journal. https://www.wsj.com/articles/tiktok-algorithm-video-investigation-11626877477.

Wang, Y., Xie, X., Fardouly, J., Vartanian, L. R., & Lei, L. (2021). The longitudinal and reciprocal relationships between selfie-related behaviors and self-objectification and appearance concerns among adolescents. New Media & Society, 23, 56–77. https://doi.org/10.1177/1461444819894346

Wells, G., Horwitz, J., & Seetharaman, D. (2021, September 14). Facebook knows Instagram is toxic for many teen girls, company documents show. The Wall Street Journal. https://www.wsj.com/articles/facebook-knows-instagram-is-toxic-for-teen-girls-company-documents-show-11631620739.

Weyzig, F. (2009). Political and economic arguments for corporate social responsibility: Analysis and a proposition regarding the CSR agenda. Journal of Business Ethics, 86, 417–428. https://doi.org/10.1007/s10551-008-9855-4

Wick, M. R., & Keel, P. K. (2020). Posting edited photos of the self: Increasing eating disorder risk or harmless behavior? International Journal of Eating Disorders, 53, 864–872. https://doi.org/10.1002/eat.23263

Wilksch, S. M., O'Shea, A., Ho, P., Byrne, S., & Wade, T. D. (2020). The relationship between social media use and disordered eating in young adolescents. *International Journal of Eating Disorders*, 53(1), 96–106. https://doi.org/10.1002/eat.23198

Zhong, R., & Frenkel, S. (2020, August). A third of TikTok's US users may be 14 or under, raising safety questions. 〈https://www.nytimes.com/2020/08/14/technology/tiktok-underage-users-ftc.html〉.

Zorthian, J. (2021, October). Washington wants to regulate Facebook's algorithm. That might be unconstitutional. *Time*. 〈[illegible]〉.

7

Digital Addiction

Hunt Allcott, Matthew Gentzkow, and Lena Song*

March 7, 2022

Abstract

Many have argued that digital technologies such as smartphones and social media are addictive. We develop an economic model of digital addiction and estimate it using a randomized experiment. Tem- ... ave persistent effects, suggesting social media are habit ... future screen time substantially reduces use, suggesting ... ggests people are inattentive to habit formation and par- ... ing at these facts through the lens of our model suggests ... social media use.

... 6, O33.

... s, self-control, temptation, naivete, commitment devices,

Addicto
Feature
cera)

... ott@microsoft.com. Gentzkow: Stanford University and NBER. ... a.song@nyu.edu. We thank Dan Acland, Matthew Levy, Peter Maxted, ... s at the Behavioral Economics Annual Meeting, the Berkeley-Chicago ... sity, Chicago Harris, Columbia Business School, Cornell, Di Tella University, the Federal Trade Commission Microeconomics Conference, Harvard, HBS, London Business School, London School of Economics, the Marketplace Innovation Workshop, Microsoft Research, MIT, the National Association for Business Economics Tech Economics Conference, the New York City Media Seminar, the New York Fed, NYU, Paris School of Economics, Princeton, Stanford Institute for Theoretical Economics, Trinity College Dublin, University of British Columbia, University College London, USC, Wharton, and Yale for helpful comments. We thank Michael Butler, Zong Huang, Zane Kashner, Uyseok Lee, Ana Carolina Paixao de Queiroz, Houda Nait El Barj, Bora Ozaltun, Ahmad Rahman, Andres Rodriguez, Eric Tang, and Sherry Yan for exceptional research assistance. We thank Chris Karr and Audacious Software for dedicated work on the Phone Dashboard app. We are grateful to the Sloan Foundation for generous support. The study was approved by Institutional Review Boards at Stanford (eProtocol #50759) and NYU (IRB-FY2020-3618). This experiment was registered in the American Economic Association Registry for randomized control trials under trial number AEARCTR-0005796; the pre-analysis plan is available from https://www.socialscienceregistry.org/trials/5796. Replication files and survey instruments are available from https://sites.google.com/site/allcott/research. Disclosures: Gentzkow does paid consulting work for Amazon, has done litigation consulting for clients including Facebook, and has been a member of the Toulouse Network for Information Technology, a research group funded by Microsoft. Both Allcott and Gentzkow are unpaid members of Facebook's 2020 Election Research Project.

Digital Addiction

Hunt Allcott, Matthew Gentzkow, and Lena Song*

March 7, 2022

Abstract

Many have argued that digital technologies such as smartphones and social media are addictive. We develop an economic model of digital addiction and estimate it using a randomized experiment. Temporary incentives to reduce social media use have persistent effects, suggesting social media are habit forming. Allowing people to set limits on their future screen time substantially reduces use, suggesting self-control problems. Additional evidence suggests people are inattentive to habit formation and partially unaware of self-control problems. Looking at these facts through the lens of our model suggests that self-control problems cause 31 percent of social media use.

JEL Codes: D12, D61, D90, D91, I31, L86, O33.

Keywords: Habit formation, projection bias, self-control, temptation, naivete, commitment devices, randomized experiments, social media.

---

*Allcott: Microsoft Research and NBER. hunt.allcott@microsoft.com. Gentzkow: Stanford University and NBER. gentzkow@stanford.edu. Song: New York University. lena.song@nyu.edu. We thank Dan Acland, Matthew Levy, Peter Maxted, Matthew Rabin, Dmitry Taubinsky, and seminar participants at the Behavioral Economics Annual Meeting, the Berkeley-Chicago Behavioral Economics Workshop, Bocconi, Boston University, Chicago Harris, Columbia Business School, Cornell, Di Tella University, the Federal Trade Commission Microeconomics Conference, Harvard, HBS, London Business School, London School of Economics, the Marketplace Innovation Workshop, Microsoft Research, MIT, the National Association for Business Economics Tech Economics Conference, the New York City Media Seminar, the New York Fed, NYU, Paris School of Economics, Princeton, Stanford Institute for Theoretical Economics, Trinity College Dublin, University of British Columbia, University College London, USC, Wharton, and Yale for helpful comments. We thank Michael Butler, Zong Huang, Zane Kashner, Uyseok Lee, Ana Carolina Paixao de Queiroz, Houda Nait El Barj, Bora Ozaltun, Ahmad Rahman, Andres Rodriguez, Eric Tang, and Sherry Yan for exceptional research assistance. We thank Chris Karr and Audacious Software for dedicated work on the Phone Dashboard app. We are grateful to the Sloan Foundation for generous support. The study was approved by Institutional Review Boards at Stanford (eProtocol #50759) and NYU (IRB-FY2020-3618). This experiment was registered in the American Economic Association Registry for randomized control trials under trial number AEARCTR-0005796; the pre-analysis plan is available from https://www.socialscienceregistry.org/trials/5796. Replication files and survey instruments are available from https://sites.google.com/site/allcott/research. Disclosures: Gentzkow does paid consulting work for Amazon, has done litigation consulting for clients including Facebook, and has been a member of the Toulouse Network for Information Technology, a research group funded by Microsoft. Both Allcott and Gentzkow are unpaid members of Facebook's 2020 Election Research Project.

1 Introduction

Digital technologies occupy a large and growing share of leisure time for people around the world. The average person with internet access spends 2.5 hours each day on social media, and there are now 3.8 billion social media users (Kemp 2020). In a 57-country survey, people now say they spend more time consuming online media than they do watching television (Zenith Media 2019). Americans check their smartphones 50 to 80 times each day (Deloitte 2018; Vox 2020; New York Post 2017).

A natural interpretation of these facts is that digital technologies provide tremendous consumer surplus. However, an increasingly popular alternative view is that habit formation and self-control problems—what we call "digital addiction"—play a substantial role. Many argue that smartphones, video games, and social media apps may be harmful and addictive in the same ways as cigarettes, drugs, or gambling (Alter 2018; Newport 2019; Eyal 2020). The World Health Organization (2018) has listed digital gaming disorder as an official medical condition. Recent experimental studies find that social media use can decrease subjective well-being (e.g. Mosquera et al. 2019; Allcott, Braghieri, Eichmeyer, and Gentzkow 2020). Figure 1 shows that social media and smartphone use are two of the top five activities that a sample of Americans think they do "too little" or "too much." Compared to the other three top activities ordered at left (exercise, retirement savings, and healthy eating), digital self-control problems have received much less attention from economists.[1]

The nature and magnitude of digital addiction matter for a number of important questions. Should people take steps to limit the amount of time they and their children spend on their smartphones and social media? What is the best way to design digital self-control tools? How can companies that make video games, social media, and smartphones best align their products with consumer welfare? Are proposed regulations such as the Social Media Addiction Reduction Technology (SMART) Act a good idea?[2]

In this paper, we formalize an economic model of digital addiction, use a randomized experiment to provide model-free evidence and estimate model parameters, and use the model to simulate the effects of habit formation and self-control problems on smartphone use. We focus on six apps that account for much of smartphone screen time and that participants report to be especially tempting: Facebook, Instagram, Twitter, Snapchat, web browsers, and YouTube. We refer to these apps as "FITSBY."

Our model follows Gruber and Kőszegi (2001), Gul and Pesendorfer (2007), Bernheim and Rangel (2004), and others in defining addiction as the combination of two key forces: habit formation and self-control problems. As in Becker and Murphy (1988), habit formation means that today's consumption increases tomorrow's demand. As in Laibson (1997) and others, self-control problems mean that people consume more today than they would have chosen for themselves in advance. These two forces are central to classic addictive goods such as cigarettes, drugs, and alcohol.

[1] Among many important examples, see Charness and Gneezy (2009) and Carrera et al. (2021) on exercise, Madrian and Shea (2001) and Carroll et al. (2009) on retirement savings, and Sadoff, Samek, and Sprenger (2020) on healthy eating.

[2] This bill, introduced in 2019 by Republican Senator Josh Hawley, proposed to prohibit the use of design features such as infinite scroll and autoplay believed to make social media more addictive, and to require companies to default users into a limit of 30 minutes per day of social media use. See Hawley (2019).

Our model allows for projection bias (Loewenstein, O'Donoghue, and Rabin 2003), where people choose as if they are inattentive to habit formation, as well as naivete about self-control problems. As in Becker and Murphy (1988), people who perceive at least some habit formation would reduce consumption if they know the price will increase in the future, while projection bias would dampen that effect. As in many other models (see Ericson and Laibson 2019), people who are at least partially aware of self-control problems might want commitment devices to restrict future consumption, and people who are at least partially unaware will underestimate future consumption.

For our experiment, we used Facebook and Instagram ads to recruit about 2,000 American adults with Android smartphones and asked them to install Phone Dashboard, an app designed for our experiment that records smartphone screen time and allows participants to set screen time limits. Participants completed four surveys at three-week intervals—a baseline (survey 1) and three follow-ups (surveys 2, 3, and 4)—that included survey measures of smartphone addiction and subjective well-being as well as predictions of future FITSBY use. Participants answered three text message survey questions per week and kept Phone Dashboard installed for six weeks after survey 4.

We independently randomized two treatments. The *bonus treatment* was a temporary subsidy of $2.50 per hour for reducing FITSBY use during the three weeks between surveys 3 and 4. We informed people whether or not they were assigned to the bonus treatment in advance, on survey 2. The *limit treatment* made available screen time limit functionality in Phone Dashboard. Participants in this group could set personalized daily time limits for each app on their phone, with changes effective the next day. These limits forced participants to stop using the relevant app and in most cases could not be immediately overridden, unlike the flexible limits in existing tools such as Android's Digital Wellbeing and iOS's Screen Time. The surveys encouraged participants to set limits in line with their self-reported ideal screen time, but doing so was entirely optional. We used multiple price lists (MPLs) to elicit participants' valuations of the bonus treatment and the limit functionality.

The bonus treatment had persistent effects that are consistent with habit formation. The bonus reduced FITSBY use by 56 minutes per day during the three weeks when the incentives were in effect, a 39 percent reduction from the control group average. In the three weeks after the incentive had ended, the bonus treatment group still used 19 minutes less per day. In the three weeks after that, they used 12 minutes less per day.

Participants correctly predict habit formation: the effects of the bonus on predicted post-incentive FITSBY use line up closely with the effects on actual use. However, in the three weeks between when the bonus was announced and when it took effect, there was only a modest (and possibly zero) anticipatory response, which is only 12 percent of what our model would predict for forward-looking habit formation without projection bias. These results are consistent with a form of projection bias in which consumers are aware of habit formation while consuming as if they are inattentive to it.[3]

[3]This distinction between awareness and attention raises interesting questions about other evidence of projection bias. For example, Busse et al. (2015) find that people are more likely to buy a convertible on sunny days. On sunny days, do people have

We also find clear evidence that people have self-control problems and are at least partly aware of them. The limit treatment reduced FITSBY screen time by 22 minutes per day (16 percent) over 12 weeks. The effects decline slightly over the course of the experiment; this decline is consistent with some loss of motivation, but the fact that the decline is slight means that the effects are unlikely to be driven by confusion or temporary novelty. Although the experiment offered no incentive to set limits, 78 percent of participants set binding limits and continued using them through the final weeks of the experiment. This far exceeds takeup of almost all commitment devices studied in the literature reviewed by Schilbach (2019, Table 1). On average, participants were willing to give up $4.20 for three weeks of access to the limit functionality, and when trading off the bonus versus a fixed payment, 24 percent said they valued the bonus more highly because they wanted to give themselves an incentive to reduce consumption. These distinct measures of commitment demand are correlated with each other and with survey measures of addiction and desire to reduce screen time.

Notwithstanding their demand for commitment, participants seem to slightly underestimate their self-control problems. The control group modestly but repeatedly underestimated their future FITSBY use in all of our surveys, even though use is fairly steady over time and we reminded them of recent past use before asking them to predict. On average, the control group underestimated next-period FITSBY use by 6.1 minutes per day, or about 4 percent.

To further evaluate whether our interventions reduced addiction in a way that participants perceive to be beneficial, we examine effects on a variety of survey outcomes. On both the main surveys and text messages, the bonus and limit treatments significantly reduced an index of smartphone addiction adapted from the psychology literature. For example, both treatment groups reported being less likely to use their phone longer than intended, use their phone to distract from anxiety or fall asleep, have difficulty putting down their phone, lose sleep from phone use, procrastinate by using their phone, and use their phone mindlessly. Both treatment groups reported improved alignment between ideal and actual screen time. The bonus treatment group also scored higher on an index of subjective well-being, with statistically significant increases in components related to concentration and avoiding distraction and statistically insignificant changes in measures of happiness, life satisfaction, anxiety, and depression. Finally, both treatments are well-targeted in the sense that effects were more positive for people who report more interest in reducing their use and who score higher on our addiction measures at baseline.

In the final section of the paper, we look at these results through the lens of our structural model. The model allows us to translate our short-run experimental estimates into effects on long-run steady state behavior, to quantify the magnitude of the effects we observe in terms of economically meaningful parameters, and to decompose the role of different behavioral forces through counterfactuals. We first estimate the model parameters by matching key moments from the experiment. We model the limit treatment as eliminating share ω of self-control problems, and for our primary estimates we conservatively assume $\omega = 1$. The estimates reflect our experimental results: substantial habit formation and self-control problems, substantial

different beliefs about future weather or how much they would drive a convertible?

projection bias, and slight naivete about self-control problems. We then evaluate how steady-state consumption would change in counterfactuals where we eliminate self-control problems. Without habit formation, a conservative estimate of the effect of self-control problems is the effect of giving people screen time limit functionality: 22 minutes per day. But habit formation amplifies the effect of self-control problems, as the increase in current consumption also increases future marginal utility. In the presence of habit formation, our primary model prediction is that eliminating self-control problems would reduce FITSBY use by 48 minutes per day, or 31 percent of baseline use. Alternative assumptions mostly imply more self-control problems, more attention to habit formation, and larger effects on use.

Our results should be interpreted with caution for several reasons. First, our experiment took place during the beginning of the coronavirus pandemic. Our survey evidence suggests that this increased screen time but did not have clear effects on the magnitude of self-control problems. Furthermore, even as the pandemic evolved over the three-month experiment, average screen time and the treatment effects of the limit were fairly stable. Second, our estimates apply to the 2,000 people who selected into our experiment, and these people are not representative of U.S. adults. When we reweight our estimates to more closely approximate national average demographic characteristics, the modeled effect of self-control problems increases. Third, our model's predictions of FITSBY use without self-control problems depend on assumptions such as linear demand and geometric decay of habit stock. Fourth, our analysis is partial equilibrium in the sense that we do not model network effects and other externalities across users. If one person's social media use increases others' use, such positive network externalities would magnify the effects of self-control problems on population-wide social media use. Finally, our surveys walked participants through a process of setting optional screen time limits that implemented their self-reported ideal screen time, and we hypothesize that simply offering time limit functionality without walking through that process would have had smaller effects.[4]

Our work builds on several existing literatures. We extend a distinguished literature documenting present focus in diverse settings including exercise, healthy eating, consumption-savings decisions, and laboratory tasks (Ericson and Laibson 2019).[5] Ours is one of a small handful of papers that estimate the parameters of a present focus model with partial naivete using field (instead of laboratory) behavior.[6] The digital self-control

[4]While Carrera et al. (2021) show that takeup of commitment devices can be driven by experimenter demand effects or decision-making noise instead of perceived self-control problems, there are three reasons why their concerns are less likely to apply to our experiment. First, while Carrera et al. (2021) studied one-time takeup of an unfamiliar commitment contract, our participants repeatedly set and continually kept screen time limits over a 12-week period. Second, we estimate even larger perceived self-control problems using participants' valuations of the bonus treatment, which leverages an alternative methodology favored by Carrera et al. (2021) as well as Acland and Levy (2012), Augenblick and Rabin (2019), Chaloupka, Levy, and White (2019), Allcott, Kim, Taubinsky, and Zinman (2021), and Strack and Taubinsky (2021). Third, unlike Carrera et al. (2021), we find strong correlations between use of screen time limits and other measures of perceived self-control problems.

[5]This includes Read and Van Leeuwen (1998), Fang and Silverman (2004), Shapiro (2005), Shui and Ausubel (2005), Ashraf, Karlan, and Yin (2006), DellaVigna and Malmendier (2006), Paserman (2008), Gine, Karlan, and Zinman (2010), Duflo, Kremer, and Robinson (2011), Acland and Levy (2012), Andreoni and Sprenger (2012a; 2012b), Augenblick, Niederle, and Sprenger (2015), Beshears et al. (2015), Goda et al. (2015), Kaur, Kremer, and Mullainathan (2015), Laibson et al. (2015), Royer, Stehr, and Sydnor (2015), Exley and Naecker (2017), Augenblick (2018), Kuchler and Pagel (2018), Toussaert (2018), Augenblick and Rabin (2019), Casaburi and Macchiavello (2019), Schilbach (2019), John (2019), Toussaert (2018), and Sadoff, Samek, and Sprenger (2020).

[6]To our knowledge, these are Allcott, Kim, Taubinsky, and Zinman (2021), Bai et al. (2018), Carrera et al. (2021), Chaloupka,

problems we study are particularly interesting because this is one of the few domains where market forces have created commitment devices, such as blockers for smartphone apps, email, and websites (Laibson 2018). Our results suggest additional unmet demand for these commitment devices.

We also extend a distinguished literature on habit formation. One set of papers documents persistent impacts of temporary interventions in settings such as academic performance, energy use, exercise, hand washing, political protest, smoking, recycling, voting, water use, and weight loss.[7] We provide evidence in an important new domain. A second set of papers tests for forward-looking habit formation using belief elicitation or advance responses to future price changes, sometimes interpreting such forward-looking behavior as support for "rational" models of addiction.[8] We estimate anticipatory responses using an experimental approach that, like the one in Hussam et al. (2019), addresses many confounds that arise in observational data (Chaloupka and Warner 1999; Gruber and Köszegi 2001; Auld and Grootendorst 2004; Rees-Jones and Rozema 2020). Furthermore, we use our model to actually estimate the magnitude of projection bias, which is important because earlier studies that reject a null hypothesis of fully myopic habit formation could still be consistent with substantial projection bias.

Finally, we extend three literatures that speak directly to digital addiction. The first literature includes theoretical papers modeling temptation in digital networks (Makarov 2011; Liu, Sockin, and Xiong 2020). The second includes experimental papers studying the effects of social media use on outcomes such as subjective well-being and academic performance.[9] The third studies the effects of digital self-control tools.[10] Hoong (2021) is particularly related, and is an important antecedent to our study. In a smaller-scale experiment, she pioneers the use of encouragement to adopt self-control tools, compares predicted and ideal use to actual use, and shows results consistent with significant self-control problems. Our paper helps to unify the previous empirical literature with a formal model of digital addiction, relatively large sample, multiple treatment arms that convincingly identify habit formation and self-control problems using several different strategies, and robust measurement of screen time and survey outcomes.

Section 2 sets up the model. Sections 3–5 detail the experimental design, data, and model-free results. Section 6 presents the model estimation strategy and parameter estimates, and Section 7 presents the modeled effects of temptation on time use.

Levy, and White (2019), and Skiba and Tobacman (2018).

[7]This includes Gerber, Green, and Shachar (2003), Charness and Gneezy (2009), Gine, Karlan, and Zinman (2010), Ferraro, Miranda, and Price (2011), John et al. (2011), Allcott and Rogers (2014), Bernedo, Ferraro, and Price (2014), Acland and Levy (2015), Royer, Stehr, and Sydnor (2015), Fujiwara, Meng, and Vogl (2016), Levitt, List, and Sadoff (2016), Beshears and Milkman (2017), Brandon et al. (2017), Carrera et al. (2018), Allcott, Braghieri, Eichmeyer, and Gentzkow (2020), Bursztyn et al. (2020), Gosnell, List, and Metcalfe (2020), and Van Soest and Vollaard (2019).

[8]This includes Chaloupka (1991), Becker, Grossman, and Murphy (1994), Gruber and Köszegi (2001), Acland and Levy (2015), Hussam et al. (2019), and Do and Jacoby (2020).

[9]This includes Sagioglu and Greitemeyer (2014), Tromholt (2016), Hunt et al. (2018), Vanman, Baker, and Tobin (2018), Mosquera et al. (2019), Allcott, Braghieri, Eichmeyer, and Gentzkow (2020), and Collis and Eggers (2019).

[10]This includes Marotta and Acquisti (2017) and Acland and Chow (2018).

2 Model

The goal of the model is to formalize the meaning of "digital addiction" and foreshadow how we identify the model parameters using our experiment.

In each period $t \leq T$, consumers choose consumption of a good x_t sold at price p_t that delivers flow utility $u_t(x_t; s_t, p_t)$. To model habit formation, utility depends on a stock s_t of past consumption that evolves according to

$$s_{t+1} = \rho(s_t + x_t), \tag{1}$$

where $\rho \in [0,1)$ captures the strength of habit formation. Habit formation captures why temporary price changes generate persistent effects in our experiment.

To model self-control problems, we follow Banerjee and Mullainathan (2010) in modeling x as a temptation good. Before period t, consumers consider period t flow utility to be $u_t(x_t; s_t, p_t)$. In period t, however, consumers choose as if period t flow utility is $u_t(x_t; s_t, p_t) + \gamma x_t$, where $\gamma \geq 0$ reflects the amount of temptation. If $\gamma > 0$, consumers choose more x_t in period t than they would choose in advance. This temptation good framework generates similar predictions to the quasi-hyperbolic model from Laibson (1997) and Gruber and Köszegi (2001), but it naturally matches our application to a single addictive good and yields simpler estimating equations where temptation is additively separable.

Consumers may misperceive temptation: before period t, consumers predict that in period t, they will consider flow utility to be $u_t(x_t; s_t, p_t) + \tilde{\gamma} x_t$. We say that consumers are fully naive if $\tilde{\gamma} = 0$, and fully sophisticated if $\tilde{\gamma} = \gamma$. Partial naivete captures why our experiment participants underestimate x_t when asked to predict in advance. Partial sophistication captures why our participants want commitment devices to change their future behavior.

Following Loewenstein, O'Donoghue, and Rabin (2003), we allow the possibility of projection bias, in which consumers choose as if to maximize a weighted average of utility given the current habit stock s_t and utility given the predicted habit stock $\tilde{s}_r$ in future period $r > t$. We let α denote the weight on the current habit stock, and thus the magnitude of projection bias. Projection bias captures why consumers in our experiment might not reduce consumption in anticipation of a known future price change. We assume that consumers are fully naive about projection bias; sophistication would introduce strategic incentives to adjust current consumption to offset future bias.[11]

Following O'Donoghue and Rabin (1999) and others, we solve for perception-perfect equilibrium strategies, where consumers maximize current utility given predictions of future behavior. Let $x_t(s_t, \gamma, \boldsymbol{p}_t)$ denote a strategy of the period-t self, which depends on habit stock, temptation, and the vector of future prices $\boldsymbol{p}_t = \{p_t, p_{t+1}, ..., p_T\}$. Let $\tilde{x}_r(s_r, \tilde{\gamma}, \boldsymbol{p}_r)$ be a consumer's *prediction*, as of period $t < r$, of her period-r

[11]Loewenstein, O'Donoghue, and Rabin (2003, page 1219) also assume naivete about projection bias, writing that "because this time inconsistency derives solely from misprediction of future utilities, it would make little sense to assume that the person is fully aware of it." We note that our formulation of projection bias is slightly different than in Loewenstein, O'Donoghue, and Rabin (2003): while their consumers' predictions of future consumption are biased due to projection bias, our consumers predict consumption accounting for habit formation, but choose as if they are inattentive to it. This matches our empirical results.

strategy. A strategy profile $(x_0^*, ..., x_T^*)$ is perception perfect if in each period t

$$x_t^*(s_t, \gamma, \boldsymbol{p}_t) = \arg\max_{x_t} u_t(x_t; s_t, p_t) + \gamma x_t + \left[\begin{array}{c} \alpha \sum_{r=t+1}^{T} \delta^{r-t} u_r(\tilde{x}_r^*(s_t, \tilde{\gamma}, \boldsymbol{p}_r); s_t, p_r) \\ +(1-\alpha) \sum_{r=t+1}^{T} \delta^{r-t} u_r(\tilde{x}_r^*(\tilde{s}_r, \tilde{\gamma}, \boldsymbol{p}_r); \tilde{s}_r, p_r) \end{array} \right], \quad (2)$$

where $\delta \leq 1$ is the discount factor.

Predicted and actual consumption differ due to naivete about temptation and projection bias and the resulting misprediction of habit stock. We assume that the equilibrium prediction $\tilde{x}_r^*(s_r, \tilde{\gamma}, \boldsymbol{p}_r)$ is the solution to equation (2) with $\alpha = 0$ and $\gamma = \tilde{\gamma}$. Predicted habit stock $\tilde{s}_r$ evolves according to $\tilde{s}_{r+1} = \rho(\tilde{s}_r + \tilde{x}_r^*(\tilde{s}_r, \tilde{\gamma}, \boldsymbol{p}_r))$.[12] The "rational" habit formation model of Becker and Murphy (1988) is the special case with $\alpha = 0$ and $\tilde{\gamma} = \gamma = 0$.

To estimate the model, we follow Becker and Murphy (1988) and Gruber and Köszegi (2001) in specializing to the case of quadratic flow utility:

$$u_t(x_t; s_t, p_t) = \frac{\eta}{2} x_t^2 + \zeta x_t s_t + \phi s_t + (\xi_t - p_t) x_t \quad (3)$$

where $\eta < 0$ measures the demand slope, ζ regulates the extent of habit formation, ϕ is the direct effect of habit stock on utility (which could be positive or negative), and ξ_t is a deterministic period-specific demand shifter. This can be microfounded by assuming that consumers have income w that they must spend in each period, and income not spent on x_t is spent on a numeraire $c_t = w - p_t x_t$ that is additively separable in u_t. In this specification, u_t is in units of dollars per period.

3 Experimental Design

3.1 Overview

Our experiment is designed to provide direct evidence on the magnitude of habit formation, perceived habit formation, temptation, and perceived temptation, as well as to identify the remaining key parameters of the quadratic model. The experiment ran from March 22 to July 26, 2020, with participants completing an intake questionnaire and four surveys. Figure 2 summarizes the experimental design, and Table 1 presents sample sizes at each step.

Between March 22 and April 8, we recruited participants using Facebook and Instagram ads. Appendix Figure A1 presents the ads. To minimize sample selection bias, the ads did not hint at our research questions or suggest that the study was related to smartphone use or social media. 3,271,165 unique users were shown one of the ads, of whom 26,101 clicked on it. This 0.8 percent click-through rate is close to the average click-through rate on Facebook ads (Irvine 2018).

[12] Since the predicted equilibrium strategy $\tilde{x}_r^*(s_r, \tilde{\gamma}, p_r)$ conditions on the state s_r inherited at time r, it will be the same when evaluated in all periods $t < r$. However, the predicted *action* in period r is not generally the same when evaluated in all periods $t < r$, as the predicted $\tilde{s}_r$ will differ by t.

Clicking on the ad took the participant to a brief screening survey, which included several background questions, the consent form, and instructions on how to install Phone Dashboard. To be eligible, participants had to be a U.S. resident between 18 and 64 years old, use an Android as their primary phone, and use only one smartphone regularly. 18,589 people satisfied these criteria, of whom 8,514 consented to participate in the study. Of these, 5,320 successfully installed Phone Dashboard and finished the intake survey.

Surveys 1–4 were administered on Sundays at three week intervals between April 12th and June 14th. We define $t = 1, 2, 3, \ldots$ to be the three-week periods beginning Monday April 13th, so period t is the three weeks immediately after survey t. For our data analysis and interventions, we want to exclude survey days, so all periods are 20 days long, from a Monday to a Saturday. Survey 1 recorded participant demographics. We describe the other survey content below.

As illustrated in Figure 2, we randomized participants into bonus and limit treatment conditions (detailed below) using a factorial design. We randomized participants to the Bonus, Bonus Control, or the Multiple Price List (MPL) group with 25, 75, and 0.2 percent probability, respectively. We independently randomized participants to the Limit or Limit Control groups with 60 and 40 percent probability, respectively. We refer to the intersection of the Bonus Control and Limit Control groups as the Control group. We balanced the randomization within eight strata defined by above- versus below-median baseline FITSBY use, *restriction index*, and *addiction index* (described below). The treatments began on survey 2.

All participants received $5 for completing the baseline survey and $25 if they completed the remaining surveys and kept Phone Dashboard installed through July 26th. Participants were also entered in a drawing for a $500 gift card, in which two winners were drawn.

As shown in Table 1, 4,038 participants completed survey 1. We dropped 1,912 of these participants from the experiment after survey 1 because they reported that they already used another app to limit their phone use (5 percent of the sample) or failed data quality checks.[13] The remaining 2,126 participants were invited to take survey 2, of whom 2,053 opened the survey and reached the point where the treatments began. Of those, 1,938 completed the study—remarkably low attrition for a 12-week study with multiple surveys.

In addition to back-loading the survey payments, several other factors contributed to our limited attrition. There were two surveys (the intake and survey 1) before the treatments began, inducing likely attriters to attrit beforehand. At the beginning of survey 2, just before the treatments began, we informed people that "anyone who drops out after this page can really damage the entire study," and offered them a choice to drop out at that moment or commit to finishing the whole study. For participants who had not yet completed each of surveys 2–4, we sent daily reminders for six days after the survey had been fielded, and after four days we began offering an additional payment for completing all remaining surveys. We also sent reminder emails to people who had failed to respond to two consecutive text messages.

[13]Participants failed data quality checks if they (i) did not to promise to "provide my best answers" on our surveys; (ii) reported having idiosyncratic bugs with Phone Dashboard; (iii) failed to answer more than two of our text message questions between survey 1 and survey 2; (iv) had a device that was incompatible with Phone Dashboard; or (v) were missing screen time data during the baseline period.

3.2 Phone Dashboard

Phone Dashboard is an Android app that was developed by a company called Audacious Software for our experiment. Appendix Figure A2 presents screenshots. Our experiment includes only Android users because a similar functionality cannot be implemented by third-party apps on iOS.

Phone Dashboard records the app that is in the foreground of a smartphone every five seconds when the screen is on; we use these data to construct our measure of consumption. It does not record the content that the user is viewing within the app. Users can see their cumulative screen time by day and by week on the Phone Dashboard home screen. This usage information was designed to be particularly useful for participants in the Bonus and Limit groups who might want to track their usage, but the Control group also used the app: the Bonus, Limit, and Control groups used Phone Dashboard for an average of 1.4, 1.5 and 1.0 minutes per day during periods 2–5.

3.3 Bonus Treatment

The bonus treatment was designed to identify projection bias (the parameter α), actual habit formation (ρ and ζ), and the curvature of utility (η). To facilitate the multiple price list (MPL) described below, survey 2 explained the bonus to all participants before telling them whether they were selected to receive it and when it would be in force. Participants were told,

If you're selected for the Screen Time Bonus, you would receive $50 for every hour you reduce your average daily FITSBY screen time below a Bonus Benchmark of [X] hours per day over the 3-week period, up to $150.

The survey then gave several examples, including:

- *If you reduce your FITSBY screen time to $[X-1] hours and 30 minutes per day over the next 3 weeks, you would receive $25.*
- *If your FITSBY screen time is above $X hours per day, you would receive $0.*

We set the Bonus Benchmark [X] as the participant's average FITSBY hours per day from period 1, rounded up to the nearest integer.

After the MPL described below, the Bonus group was informed that they had been randomly selected to receive the bonus for screen time reductions during period 3—i.e., starting in three weeks. The Bonus Control group was informed that they would not receive the bonus. To ensure that participants understood, each participant had to answer a question by correctly indicating their bonus treatment condition before advancing. We also sent three text messages reminders to the Bonus group during period 2, which read "Don't forget, we'll pay you $50 for every hour you reduce your average daily screen time between May 24 and June 14. There is no bonus for changing your screen time before then." People were asked to respond to the text message to confirm that they had read it. Survey 3 included an additional reminder for the Bonus treatment group. While we received substantial feedback on the surveys and many emails from our 2,000

participants during the study and our earlier pilots, none of these interactions suggested confusion about the timing of the bonus.

The Bonus group's anticipatory response to the bonus in period 2 (before the incentive was in effect) provides information about the magnitude of projection bias α. The contemporaneous response in period 3 (while the incentive was in effect) provides information about the price response parameter η. The long-term effects in periods 4 and 5 (after the incentive had ended) provides information about the magnitude and decay of habit (ζ and ρ).

3.4 Limit Treatment

The limit treatment was designed to understand self-control problems and help identify the temptation parameter γ. The Limit treatment group was given access to functionality in Phone Dashboard that allows users to set daily time limits for each app on their phone; see Appendix Figure A2 for screenshots. Any changes to the limits take effect the next day. Phone Dashboard serves five-minute and one-minute push notifications as an app's daily time limit approaches. When the limit arrives, users can "snooze" their limit and get an additional amount of time that they specify—but starting only after a delay. Within the Limit group, we randomly assigned participants with equal probability to delays of 0, 2, 5, or 20 minutes or a condition where the ability to snooze was disabled. To keep the scope of this paper manageable, we focus only on the comparison between the Limit and Limit Control groups; we plan to study the variation in snooze delays in a separate paper. To reduce attrition and uninstallation, Phone Dashboard also allows people to permanently opt out of the limits; about 4 percent of the Limit group did so.

The Limit group was first given access to the Phone Dashboard limit functionality on survey 2, after the Screen Time Bonus multiple price list described below, and they retained access to the feature for the duration of the experiment. To introduce the limits, we first gave participants instructions on how to set daily app usage limits for themselves. The survey then asked participants what time limits they would like to set for themselves on each FITSBY app over the next three weeks. We then asked participants to update their Phone Dashboard app, which activated the limit functionality, and encouraged them to set the limits they had reported a moment earlier. The Limit Control group was never told about limits and continued to have a version of Phone Dashboard that did not have the limit functionality.

In the analysis below, we interpret use of the limits as evidence of perceived self-control problems ($\tilde{\gamma} > 0$).

3.5 Bonus and Limit Valuations

We used incentive-compatible multiple price list mechanisms to elicit valuations of the Screen Time Bonus and the limit functionality. Because both the bonus and the limit functionality reduce future social media use, these valuations help identify perceived temptation $\tilde{\gamma}$.

All multiple price lists included a table with a series of choices between "Option A" and "Option B"

in separate rows. Option B was the same in each row, while Option A included an amount of money that decreased monotonically from top to bottom. Participants would typically choose Option A at the top and Option B at the bottom, and we infer their valuation of Option B from the row where they switch. To encourage valid answers, participants who did not switch between Option A and Option B exactly once were alerted to this fact and given a chance to change their answers. All MPLs were incentivized, as described below. To help participants become familiar with MPLs, survey 1 included an incentivized practice MPL that asked participants to choose between receiving different survey completion payments at different times.

Our approach to valuing the Screen Time Bonus builds on Allcott, Kim, Taubinsky, and Zinman (2021) and Carrera et al. (2021). Survey 2 informed participants of their average daily FITSBY screen time over the past three weeks and asked them to predict their screen time over the next three weeks. The survey then introduced the Screen Time Bonus and asked participants to predict how much they would reduce their FITSBY screen time relative to their original prediction if they were selected for the bonus.

After these two predictions, we asked participants to make a hypothetical choice between the Screen Time Bonus and a payment equal to their expected earnings from the bonus. The survey described potential considerations as follows:

- *You might prefer $[expected earnings] instead of the Screen Time Bonus if you don't want any pressure to reduce your screen time.*
- *You might prefer the Screen Time Bonus instead of $[expected earnings] if you want to give yourself extra incentive to use your phone less.*

Participants then completed an MPL where Option B was receiving the Screen Time Bonus, and Option A was receiving a payment ranging from $150 to $0.

To make the MPL incentive compatible, participants were told, "Last week, the computer randomly selected some participants to receive what they choose on the multiple price list below, and also randomly selected one of the rows to be 'the question that counts.' If you were randomly selected to participate, you will be paid based on what you choose in that row." 0.2 percent of participants were randomly assigned to the MPL group that received what they chose on a randomly selected row.

On survey 3, the Limit group completed an MPL that elicited valuations of the Phone Dashboard limit functions. Option B was retaining access to the Phone Dashboard limit functions, and Option A was having those functions disabled for the following three weeks in exchange for a dollar payment that ranged from $20 to -$1. The MPL group received what they chose on a randomly selected row.

3.6 Predicted Use

At the end of surveys 2, 3 and 4, we elicited predictions of future FITSBY use. These predictions help identify the degree of naivete or sophistication about temptation—the difference between γ and $\tilde{\gamma}$.

Before each elicitation, we told each participant their average FITSBY screen time over the previous three weeks. Surveys 2 and 3 also reminded the Bonus and Limit groups about the bonus and limits. Survey

2 then elicited predictions of FITSBY screen time for the next three weeks (period 2), the three weeks after that (period 3), and the three weeks after that (period 4). Survey 3 elicited separate predictions for periods 3, 4, and 5. Survey 4 elicited separate predictions for periods 4 and 5.

Predictions were incentivized. Survey 2 told participants, "Answer carefully, because you might earn a Prediction Reward. After the study ends, we will pick a prediction question at random and check how close your prediction is. If your predicted daily screen time is within 15 minutes of your actual screen time, we will pay you an additional $X." We randomized the prediction reward X to be $1 or $5, each with 50 percent probability.

3.7 Survey Outcome Variables

Surveys 1, 3, and 4 asked questions designed to measure participants' perceptions of their addiction and subjective well-being (SWB). For the nine weeks between survey 1 and survey 4, we also sent three text messages per week with a subset of questions that we thought were important to ask in real time instead of retrospectively. Using these questions, we construct five pre-specified outcome variables. Appendix A.1 presents details on the survey questions.

Ideal use change. The survey said,

Some people say they use their smartphone too much and ideally would use it less. Other people are happy with their usage or would ideally use it more. How do you feel about your smartphone use over the past 3 weeks?

- *I use my smartphone too much.*
- *I use my smartphone the right amount.*
- *I use my smartphone too little.*

For people who said they used their smartphone "too much" or "too little," we then asked, *Relative to your actual use over the past 3 weeks, by how much would you ideally have [reduced/increased] your smartphone use?* The *ideal use change* variable is the answer to this question, in percent.

Addiction scale. Our addiction scale is a battery of 16 questions modified from two well-established survey scales, the Mobile Phone Problem Use Scale (Bianchi and Phillips 2005) and the Bergen Facebook Addiction Scale (Andreassen et al. 2012). The questions attempt to measure the six core components of addition identified in the addiction literature: salience, tolerance, mood modification, relapse, withdrawal, and conflict (Griffiths 2005).

The survey asked, *In the past three weeks, how often have you ...*, with a matrix of 16 questions, such as

- *used your phone longer than intended?*
- *felt anxious when you don't have your phone?*

- *lost sleep due to using your phone late at night?*

Possible answers were Never, Rarely, Sometimes, Often, and Always, which we coded as 0, 0.25, 0.5, 0.75, and 1, respectively. *Addiction scale* is the sum of these numerical scores for the 16 questions.

SMS addiction scale. The SMS addiction scale includes shortened versions of nine questions from the addiction scale. Examples include:

- *In the past day, did you feel like you had an easy time controlling your screen time?*
- *In the past day, did you use your phone mindlessly?*
- *When you woke up today, did you immediately check social media, text messages, or email?*

People were instructed to text back their answers on a scale from 1 (not at all) to 10 (definitely). *SMS addiction scale* is the sum of these scores for the nine questions.

Phone makes life better. The survey asked, *To what extent do you think your smartphone use makes your life better or worse?* Responses were on a scale from -5 ("Makes my life worse") through 0 ("Neutral") to +5 ("Makes my life better").

Subjective well-being. We use standard measures from the subjective well-being literature, mostly following the measures from our own earlier work (Allcott, Braghieri, Eichmeyer, and Gentzkow 2020). The survey asked,

Please tell us the extent to which you agree or disagree with each of the following statements. ***Over the last three weeks,*** with a matrix of seven questions:

- *... I was a happy person*
- *... I was satisfied with my life*
- *... I felt anxious*
- *... I felt depressed*
- *... I could concentrate on what I was doing*
- *... I was easily distracted*
- *... I slept well*

Possible answers were on a seven-point scale from "strongly disagree" through "neutral" to "strongly agree," which were coded as -1, -2/3, -1/3, 0, 1/3, 2/3, and 1, respectively. The variable *subjective well-being* is the sum of these numerical scores for the seven questions, after reversing *anxious*, *depressed*, and *easily distracted* so that more positive reflects better subjective well-being.

Indices. We define the *survey index* to be the sum of the five survey outcome variables described above, weighted by the baseline inverse covariance matrix as described by Anderson (2008). When presenting results and constructing this index, we orient the variables so that more positive values imply normatively better outcomes. Thus, we multiply *addiction scale* and *SMS addiction scale* by (-1).

We define the *restriction index* to be the sum of *interest in limits* (with the four categorical answers coded as 0, 1, 2, and 3) and *ideal use change*, after normalizing each into standard deviation units. We define the *addiction index* to be the sum of *addiction scale* and *phone makes life better* after normalizing each into standard deviation units. We use these two indices for stratified randomization and as moderators when testing for heterogeneous treatment effects.

3.8 Pre-Analysis Plan

We submitted our pre-analysis plan (PAP) on May 4th, the day that post-treatment data collection began. The PAP specified (i) the equation for treatment effect estimation (equation 4 below); (ii) the construction of the survey outcome variables and indices described in Section 3.7, the *limit tightness* variable, and the winsorization of predicted FITSBY use; and (iii) the analysis of heterogeneous treatment effects by splitting the sample on above- versus below-median values of six moderators: education, age, gender, baseline FITSBY use, *restriction index*, and *addiction index*. The PAP also included shells of Tables 1, 2, and A1–A3, as well as Figures 1–6, A1–A4, A8, and A28–A34.

We deviate from the PAP in five ways. First, the bottom left panel of Figure 3 includes results from each addiction scale question, whereas the PAP figure shell presented the sum across all questions. Second, we clarify that our analysis sample includes only the balanced panel of people who completed the study. Results are essentially identical if we use an unbalanced panel that includes data from attriters before they attritted, but the balanced panel is helpful in ensuring that our habit formation results are not spuriously driven by attrition. Third, three figures from the PAP are not included here, as we plan to study them in a separate paper. Fourth, Figure 6 includes predicted FITSBY use from all surveys before period t, whereas the PAP figure shell presented predictions from only the survey immediately before period t. Fifth, we use equation (4) for subgroup analysis, whereas the PAP specified that we would use an instrumental variables regression. We present the pre-specified instrumental variables estimates in Appendix D.4. The results are similar, and we decided that equation (4) was simpler.

4 Data

The analysis sample for all results reported below is the balanced panel of 1,933 participants who were assigned to either Bonus or Bonus Control (not the MPL group), completed all four surveys, and kept Phone Dashboard installed until the end of the study on July 26. This group's attrition rate after being informed of treatment was $(1 - 1{,}933/2{,}048) \times 100\% \approx 5.6$ percent. Attrition rates and observable characteristics are balanced across the bonus and limit treatment conditions; see Appendix Tables A1 and A2.

Table 2 quantifies the representativeness of our analysis sample on observables, by comparing their demographics to the U.S. adult population. Our sample is more educated, more heavily female, younger, and slightly lower-income than the U.S. population. We estimate an alternative specification of our structural model with sample weights to adjust for these observable differences.

Table 2 also shows that the average participant had 333 minutes per day of screen time during the baseline period, of which 153 minutes (46 percent) was on FITSBY apps. Different sources report very different estimates of average social media use and smartphone screen time for U.S. adults, so we do not report nationwide averages in the table. Kemp (2020) reports that internet users in the U.S. and worldwide, respectively, spend an average of 123 and 144 minutes per day on social media, mostly on mobile devices. Wurmser (2020) and Brown (2019) report national averages of 186 and 324 minutes of total smartphone screen time per day, respectively. The comparisons suggest that the heavy use in our sample may not be far from the national average.

During the baseline period, the average participant used Facebook, browsers, YouTube, Instagram, Snapchat, and Twitter for 69, 44, 23, 24, 15, and 15 minutes per day, respectively; see Appendix Figure A3. Appendix Figure A4 presents the distribution of baseline FITSBY use. Appendix Table A3 presents descriptive statistics for the survey outcome variables.

5 Model-Free Results

5.1 Treatment Effect Estimating Equation

To estimate treatment effects, define Y_{it} as an outcome for participant i for period t. Y_{it} could represent a survey outcome variable measured on survey $t \in \{3,4\}$, or period t FITSBY use. Define L_i and B_i as Limit and Bonus group indicators. Define $\boldsymbol{X}_{i1}$ as a vector of baseline covariates: baseline FITSBY use and, if and only if Y is a survey outcome variable, the baseline value Y_{i1} and the baseline value of *survey index*. Define ν_{it} as a vector of the eight randomization stratum indicators, allowing separate coefficients for each period t. We estimate the effects of the limit and bonus treatments using the following regression:

$$Y_{it} = \tau_t^B B_i + \tau_t^L L_i + \beta_t \boldsymbol{X}_{i1} + \nu_{it} + \varepsilon_{it}. \quad (4)$$

When combining data across multiple periods, we cluster standard errors by participant.

5.2 Baseline Qualitative Evidence

Figure 3 presents qualitative evidence on digital addiction from the baseline survey. The top two panels present the variables in the *restriction index*. The top left panel shows that 23 percent of people reported being "moderately" or "very" interested in setting time use limits on their smartphone apps, while 34 percent reported being "not at all" interested. The top right panel presents the distribution of responses to the *ideal use change* question. 42 percent of people said that they used their smartphone the right amount over the

past three weeks, and only 0.5 percent said that they used it too little. Among people who said they used their smartphone too much, the average ideal reduction was 34 percent.

Survey 1 also asked people to report their ideal use change for specific apps or categories. FITSBY, games, video streaming, and messaging are the nine apps on which people want to reduce screen time the most; see Appendix Figure A8. Facebook is by far the most tempting app: the average participant would ideally reduce Facebook use by 22 percent. The average participant did not want to change their use of email, news, and maps and wanted to slightly increase use of phone, music, and podcast apps.

The bottom two panels present the variables in the *addiction index*. The bottom left panel presents the share of participants who responded "often" or "always" on each question in the *addiction scale*. The top seven questions capture three components of moderate addictions (salience, tolerance, and mood modification); 33 percent of participants often or always experience each of these, and 84 percent often or always experience at least one. The bottom nine questions capture three components of more severe addictions (relapse, withdrawal, or conflict); 11 percent of participants often or always experience each of these, and 41 percent often or always experience at least one. The bottom right panel shows that while most people think that their smartphone use makes their life better, 19 percent think that it makes their life worse. Taken together, these results suggest substantial heterogeneity: many people report experiences consistent with addiction, while many others do not.

Our experiment took place during the coronavirus pandemic, which significantly disrupted people's daily routines. To understand how this might affect our results, we included several baseline survey questions, which we report in Appendix C. 78 percent of people reported having more free time as a result of the pandemic, and 88 percent of people reported that the pandemic had increased their phone use. However, it is not clear that the pandemic affected the extent of self-control problems: the means and distributions of key qualitative measures of addiction that we also asked for 2019, *ideal use change* and *phone makes life better*, were statistically different but economically similar. *Ideal use change* is closer to zero in 2020 compared to in 2019, suggesting less perceived self-control problems, but *phone makes life better* is also less positive, suggesting more perceived self-control problems.

5.3 Bonus Treatment and Habit Formation

The darker coefficients in Figure 4 present the effect of the bonus on FITSBY use, estimated using equation (4). Recall that the bonus provides an incentive to reduce FITSBY use in period 3, but we informed participants about whether or not they were offered the bonus at the beginning of period 2. The incentive is \$50 per *average* hour measured over the 20-day period, or \$2.50 per hour of consumption.

In period 3 (while the incentive was in effect), the Bonus group reduced FITSBY use by 56 minutes per day, or 39 percent relative to the Control group. This is a striking price response: it implies that participants value a substantial share of smartphone FITSBY use at less than \$2.50 per hour.

In periods 4 and 5 (after the incentive had ended), the Bonus group still reduced FITSBY use by 19 and 12 minutes per day, respectively. This persistent effect suggests substantial habit formation, implying $\zeta > 0$

in our model. The decay of the effect in period 5 relative to period 4 provides information about the habit stock decay parameter ρ in our model.

In period 2 (before the incentive was in effect), the Bonus group reduced FITSBY use by 5.1 minutes per day, which is marginally statistically significant. This is consistent with the model's prediction that a consumer who perceives habit formation should reduce period 2 consumption in order to reduce period 3 marginal utility, which makes it easier to reduce period 3 consumption in response to the financial incentive. However, additional evidence suggests some caution about interpreting the period 2 effect as forward-looking habit formation. Appendix Figures A9 and A10 break out the period 2 effect separately by day and week, showing that it loads mostly on the first half of the period. If anything, forward-looking habit formation would predict the opposite pattern, with larger anticipatory effects closer to the beginning of the incentive period. Possible explanations include intertemporal substitution, a temporary idiosyncratic effect, and the salience of the bonus after its introduction on survey 2.[14]

5.4 Limit Treatment and Temptation

The Limit group made extensive use of the limit functionality. To summarize the stringency of time limits, we define the variable *limit tightness* to be the amount by which a user's limits would have hypothetically reduced screen time if applied to their baseline use.[15] *Limit tightness* equals zero (instead of missing) for an app if the participant doesn't have the app or doesn't set a limit, so this variable speaks to what apps contribute the most to aggregate temptation. About 89 percent of the Limit group had positive *limit tightness* at some point during the experiment, suggesting that they set binding screen time limits, and 78 had positive *limit tightness* in period 5, meaning that they kept those limits for more than three weeks after the final survey. Participants most wanted to restrict Facebook, web browsers, YouTube, and Instagram: *limit tightness* averaged 20, 10, 8, and 6 minutes per day on those apps, respectively, across periods 2–5. Across all apps, the Limit group's average *limit tightness* was 53 minutes per day. See Appendix Figures A11 and A12 for details.

The lighter coefficients on Figure 4 present the effect of the limit on FITSBY use. These actual effects are smaller than the *limit tightness* values in the previous paragraph primarily because users snooze the

[14] Although we stratified randomization on period 1 FITSBY use and also control for period 1 use when estimating equation (4), some idiosyncratic factor could temporarily affect consumption in Bonus versus Bonus Control at the beginning of period 2. Some evidence supports this possibility: Appendix Figure A9 shows that consumption is slightly lower in the Bonus group compared to Bonus Control in the final 11 days of period 1. Salience could also play a role, although as described in Section 3.3, we took many steps to eliminate confusion about the timing of the bonus incentive period, and participants likely would have emailed our team if they were confused.

[15] Specifically, define x_{iadt} as the screen time of person i on app a on day d in period t. Define h_{iat} as the average screen time limit in place in period t, and define $N_{d\in t=1}$ as the number of days in the baseline period. *Limit tightness* is

$$H_{iat} = \frac{1}{N_{d\in t=1}} \sum_{d\in t=1} \max\{0, x_{iad1} - h_{iat}\}. \tag{5}$$

If the daily limit h_{iat} would not have been binding in baseline day d, the max function returns 0. If h_{iat} would have been binding in day d, then the max function returns the excess screen time on that day. We aggregate over apps to construct user-level limit tightness $H_{it} = \sum_a H_{iat}$.

limits. Access to the limit functionality reduced use in periods 2–5 by an average of 22 minutes per day, or 16 percent relative to the Control group. The effects attenuate only slightly as the experiment continues, and the effect is still 19 minutes per day in the last week of period 5. This is notable because while surveys 2 and 3 walked people through a limit-setting process and survey 4 included an optional review of the limits, the end of period 5 is nine weeks after survey 3 and six weeks after survey 4. These continuing effects suggest that while motivation might decrease over time, use of the limits is not primarily driven by confusion or temporary novelty. Furthermore, Appendix D.1 shows that *limit tightness* is correlated in expected ways with bonus and limit valuations and with survey measures of addiction and desire to reduce screen time. This evidence consistently points toward perceived self-control problems, implying $\tilde{\gamma} > 0$ in our model.

When we add the interaction between Bonus and Limit group indicators to equation (4), the main effects are similar and the interaction terms are not statistically significant; see Appendix Figure A13.

5.5 Substitution

Figure 5 presents usage effects of the bonus (in period 3 only) and the limit (across periods 2–5) separately by app. Among the FITSBY apps, Facebook sees the largest reductions, followed by web browsers, YouTube, Instagram, Twitter, and Snapchat. The effects on other apps (the right-most coefficients) provide evidence on the extent to which participants substituted FITSBY time to alternative apps. The bonus has no statistically detectable effect on use of other apps in period 3, and the confidence intervals rule out any substantial substitution relative to the 56 minutes per day reduction in FITSBY use. The limit induces substitution of 12 minutes per day, so that roughly half of the FITSBY screen time that the limit eliminates moves to other apps where people had been less likely to set limits.

One important limitation is that we cannot directly monitor FITSBY use on devices other than the participant's smartphone. We screened out potential participants who reported using more than one smartphone regularly, but our remaining participants may still have used desktops, tablets, or other devices. To provide some evidence on this substitution, survey 4 asked participants to estimate their FITSBY use on other devices in period 3 compared to the three weeks before they joined the study. The results, shown in Appendix Figure A14, imply that the limit increased FITSBY use on other devices by a marginally significant 4.2 minutes per day. The bonus *reduced* the amount of time they spent on FITSBY on other devices by 8.1 minutes per day, suggesting that time on other devices was a mild complement in this case.

The differences in substitution induced by the bonus versus limit are notable. In a simple model where other apps and devices are either complements or substitutes for smartphone FITSBY use, the substitution effects described above might have the same sign for both the bonus and limit and might be in proportion to their direct effects on smartphone FITSBY use. In contrast, a much smaller share of the effect on FITSBY use is substituted to other smartphone apps for the bonus compared to the limit, and the self-reported effects on FITSBY use on other devices have opposite signs for the bonus versus the limit. This is an interesting result to understand in future work.

5.6 Predicted versus Actual Use

Figure 6 presents predicted and actual FITSBY use in the Control condition, where participants had neither the bonus nor the limit functionality. As specified in our pre-analysis plan, we winsorize predicted use at no more than 60 minutes per day more or less than actual use in the corresponding period. Within each period, the left-most spike is actual average use. The spikes to the right are average predictions. The point estimates show that people consistently underestimate their use in all future periods, even though actual use is fairly stable throughout the experiment and the surveys had reminded them of their past use before eliciting predictions. This is consistent with naivete, implying $\tilde{\gamma} < \gamma$ in our model.

Figure 7 presents predicted versus actual habit formation. Within each period, the left-most point is the treatment effect of the bonus on actual use, reproduced from Figure 4. Recall that before the multiple price list for the Screen Time Bonus on survey 2, we asked people to report the percent by which they thought the bonus would reduce their FITSBY use. Their estimates (translated into minutes using their status quo predictions) are almost exactly correct on average: 52 minutes per day. Then on survey 3, we asked people to predict their use in future periods. Figure 7 also presents treatment effects of the bonus on predicted use, estimated from equation (4). The figure shows that people correctly predict that the bonus will reduce their consumption in period 3 and that this reduction will persist even after the incentive is no longer in effect. If anything, comparing the time path of actual versus predicted effects suggests that people overestimate the extent of habit formation. Overall, these results suggest that people are well aware of habit formation.

Appendix D.1 presents additional results that validate that the usage predictions are meaningful. Predicted use lines up well with actual use, and the higher ($5 instead of $1) prediction accuracy reward slightly reduces the absolute value of the prediction error but has tightly estimated zero effects on predicted use, actual use, and the level of the prediction error.

5.7 Bonus and Limit Valuations

On the survey 3 multiple price list, the average Limit group participant was willing to give up a $4.20 fixed payment for three weeks of access to the limit functionality. About 58 percent of participants were willing to give up at least some money for the limits, and 20 percent were willing to give up more than $10; see Appendix Figure A17. This willingness to pay for a commitment device is consistent with perceived self-control problems ($\tilde{\gamma} > 0$) and unmet market demand for digital self-control tools.

On the survey 2 multiple price list, people who perceive self-control problems should prefer the Screen Time Bonus over higher fixed payments, as the incentive helps bring future use in line with current preferences. We show in Appendix E.5 that participants' average valuation of the bonus is consistent with perceived self-control problems ($\tilde{\gamma} > 0$).

Appendix D.1 presents additional results that validate that the MPL responses are meaningful. First, participants' valuations of the bonus are correlated with the amount of money they could expect to earn. Second, the bonus and limit valuations are correlated with each other and with *limit tightness*, *ideal use*

change, *addiction scale*, *SMS addiction scale*, and other variables in expected ways. Third, after the bonus MPL, we asked people to "select the statement that best describes your thinking when trading off the Screen Time Bonus against the fixed payment." 24 percent responded that "I wanted to give myself an incentive to use my phone less over the next three weeks, even though it might result in a smaller payment," and this group had a higher average valuation.

5.8 Effects on Survey Outcomes

Figure 8 presents the effects of the bonus and limit treatments on the survey outcomes described in Section 3.7. The outcome variables are signed so more positive effects always correspond to less addiction and/or higher subjective well-being. Following our pre-analysis plan, when estimating effects on survey outcomes, we constrain the limit effect to be the same for surveys 3 and 4 (because we correctly anticipated similar "first stage" effects on FITSBY use in both periods 2 and 3) and we report the bonus effect only for survey 4 (because we correctly anticipated negligible "first stage" effects on FITSBY use in period 2).[16]

Figure 8 shows that both interventions significantly reduced self-reported measures of addiction. Appendix Table A6 presents coefficient estimates and p-values. The bonus effect is larger than the limit effect for five of the six variables, consistent with the bonus's larger effects on FITSBY use. The bonus decreased *ideal use change* by 0.41 standard deviations (about 9 percentage points), while the limit decreased it by 0.23 standard deviations (about 5 percentage points). Both interventions reduced *addiction scale* and *SMS addiction scale* by 0.08 to 0.16 standard deviations, or about 0.21–0.44 points on the 16-point *addiction scale*. Both interventions statistically significantly reduced the chance that people reported using their smartphones to relax to go to sleep, losing sleep from use, using longer than intended, using to distract from anxiety, having difficulty putting down their phone, using mindlessly, and other specific measures from the addiction scales; see Appendix Figures A23 and A24. The limit treatment statistically significantly increased the extent to which people thought their smartphone use made their life better, while the bonus did not.

The bonus and limit treatments increased subjective well-being (SWB) by 0.09 standard deviations ($p \approx 0.026$) and 0.04 standard deviations ($p \approx 0.18$) respectively. The sharpened False Discovery Rate-adjusted p-values (see Benjamini and Hochberg 1995) are 0.09 and 0.24, respectively. These SWB effects appear to be driven particularly by improved concentration and reduced distraction; see Appendix Figure A25. The effects of the bonus and limit on happiness, life satisfaction, depression, and anxiety are individually and collectively insignificant, while the effects of the bonus (but not the limit) on concentration, distraction, and sleep quality are collectively significant. Both interventions improved *survey index*, the inverse covariance-weighted average of the five survey outcome variables, by about 0.2 standard deviations.

One point of comparison for the SWB effects is Allcott, Braghieri, Eichmeyer, and Gentzkow (2020). They find that deactivating subjects' Facebook accounts for a four week period increased an index of SWB

[16] Appendix Figure A22 presents the treatment effects on survey outcomes separately for surveys 3 and 4. The limit effects on surveys 3 and 4 are statistically indistinguishable. Although the bonus did not substantially affect consumption in period 2, the Bonus group reported more ideal use reduction and more addiction on survey 3. One potential explanation is that the Bonus group hoped to reduce FITSBY use in anticipation of the period 3 incentive, and these survey responses reflect their failure to do so.

by a statistically significant 0.09 standard deviations. Although the two interventions had similar effects on time use—deactivation in Allcott, Braghieri, Eichmeyer, and Gentzkow (2020) reduced Facebook use by 60 minutes per day for 27 days, while our Screen Time Bonus reduced FITSBY use by 56 minutes per day for 20 days—they differed on a number of dimensions including the apps affected and the time period in which the study took place.

Appendix Figure A26 presents effects on *survey index* in subgroups with above- and below-median values of our six pre-specified moderators. There is little heterogeneity with respect to the first four moderators, other than that the limit seems to have larger effects on women. However, the effects of both interventions are 2–3 times larger for people with above-median baseline values of *restriction index*, which measures interest in restricting smartphone time use, and *addiction index*. This implies that the interventions are well-targeted: they have larger effects for people who report wanting and needing them the most. Consistent with this, point estimates suggest that the bonus and limit both have larger effects on FITSBY use for people with higher *restriction index* and *addiction index*, although the differences are not as significant; see Appendix Figure A27. This targeting result need not have been the case: for example, it could have been that more addicted people were less likely to feel that the limit functionality worked well for them.

6 Estimating the Model

6.1 Setup

We now turn to our model to simulate the effect of temptation on steady-state FITSBY use. In the model from Section 2, temptation and habit formation interact, because the current consumption increase caused by temptation also increases future consumption. The long-run effect of temptation could therefore be different than any effects identified during the experiment. In this section, we estimate the model's structural parameters. In the next section, we simulate steady-state FITSBY use with counterfactual self-control and habit formation parameters.

We estimate the model using indirect inference: we derive equations that characterize how a consumer from our model would behave in our experiment, and we solve for the structural parameters consistent with the data. In our baseline estimates, we assume that all parameters other than ξ are homogeneous across consumers, although we relax this assumption in an extension that allows heterogeneity in temptation γ and perceived temptation $\tilde{\gamma}$.

In describing the estimation strategy, we focus on a "restricted model" where we set the anticipatory bonus effect τ_2^B to zero. This implies full projection bias ($\alpha = 1$), and thus that consumption decisions maximize current-period flow utility with no dynamic considerations. This substantially simplifies the exposition and, as we will show, has little impact on the results. Appendix E presents our "unrestricted model," in which we use the empirical τ_2^B and allow partial projection bias.

In the restricted model, consumers maximize current-period flow utility from equation (3), giving equi-

librium consumption

$$x_t^*(s_t, \gamma, \boldsymbol{p}_t) = \frac{\zeta s_t + \xi_t - p_t + \gamma}{-\eta}. \tag{6}$$

We define $\lambda := \frac{\partial x_t^*}{\partial s_t}$ as the effect of habit stock on consumption; $\lambda = -\zeta/\eta$ in the restricted model. In a steady state with constant s, ξ, and p, we must have $s_{ss} = \rho(s_{ss} + x_{ss})$, and thus $s_{ss} = \frac{\rho}{1-\rho} x_{ss}$.

We model the Screen Time Bonus as a price $p^B = \$2.50$ per hour in period 3 plus a fixed payment.[17] We model the limit functionality as an intervention that elimates share ω of perceived and actual temptation. We conservatively assume $\omega = 1$ in our primary estimates, and we consider alternative assumptions below. We assume that when predicting period t consumption on the survey at the beginning of period t, consumers use perceived temptation $\tilde{\gamma}$ but are aware of projection bias, so the prediction is denoted $x_t^*(s_t, \tilde{\gamma}, \boldsymbol{p}_t)$.

Figure 9 illustrates temptation, naivete, and our identification strategies. The three demand curves are desired demand $x_t^*(s_t, 0, \boldsymbol{p}_t)$ according to preferences before period t, predicted demand $x_t^*(s_t, \tilde{\gamma}, \boldsymbol{p}_t)$ as of survey t, and actual demand $x_t^*(s_t, \gamma, \boldsymbol{p}_t)$. The actual equilibrium at $p = 0$ is point L, and the predicted equilibrium is at point C, so the distance CL is Control group misprediction $m^C := x_t^*(s_t, \gamma, \boldsymbol{p}_t) - x_t^*(s_t, \tilde{\gamma}, \boldsymbol{p}_t)$. The bonus moves the equilibrium to point J in period 3, so the contemporaneous bonus effect τ_3^B is the distance JK. The limit treatment moves the equilibrium to point G, so the limit treatment effect τ^L is the distance GL.

6.2 Estimating Equations

Unlike many applications of indirect inference, we derive equations that allow us to directly solve for the model parameters, so we do not need to use an optimization routine to search for parameters that fit the data. We estimate the parameters in stages, as parameters estimated in the first few equations are used as inputs to subsequent equations. We estimate confidence intervals by bootstrapping. Appendix G presents formal derivations and additional details.

Habit Formation

We first estimate ρ from the decay of the bonus treatment effects. Taking the expectations over ξ in the Bonus and Bonus Control groups, we can write the average treatment effect of the bonus on period 4 consumption as the result of the decayed period 3 effect. Similarly, the average treatment effect in period 5 results from the cumulative decayed effects from periods 3 and 4:

[17] Modeling the bonus as a linear price simplifies the model substantially, although it is an approximation: 13 percent of the Bonus group hit the \$150 payment limit because they reduced period 3 FITSBY use by more than three hours per day relative to their Bonus Benchmark, and 3.5 percent used more than their Bonus Benchmark. These two subgroups in practice faced zero subsidy for marginal screen time reductions, although they may not have known that.

$$\tau_4^B = \lambda \rho \tau_3^B \tag{7}$$
$$\tau_5^B = \lambda \left(\rho \tau_4^B + \rho^2 \tau_3^B\right). \tag{8}$$

Dividing those two equations gives

$$\rho = \frac{\tau_5^B}{\tau_4^B} - \frac{\tau_4^B}{\tau_3^B}. \tag{9}$$

This equation shows that if the bonus effect is more persistent between periods 4 and 5, we infer that habit stock is more persistent (a larger ρ).

In the unrestricted model in Appendix E, we also estimate λ, because it is useful in estimating the other parameters. To provide a comparison, we also estimate λ in the restricted model by rearranging equation (7) and inserting the ρ from equation (9): $\lambda = \frac{\tau_4^B}{\rho \tau_3^B}$.

Price Response and Habit Stock Effect on Marginal Utility

After estimating ρ, we estimate η and ζ from the magnitude and decay of the bonus treatment effects. For each of periods 3 and 4, we take the expectations over ξ of equilibrium consumption in the Bonus and Bonus Control groups, difference the two, and rearrange, giving

$$\eta = \frac{p^B}{\tau_3^B} \tag{10}$$
$$\zeta = \frac{-\eta \tau_4^B}{\rho \tau_3^B}. \tag{11}$$

Figure 9 illustrates the first equation: the inverse demand slope η is just the ratio of p^B (the vertical distance KL) to τ_3^B (the horizontal distance JK). The second equation shows that if the bonus effect is more persistent between periods 3 and 4, we infer that habit stock has a larger effect on marginal utility (a higher ζ).

Naivete about Temptation

Next, we estimate naivete about temptation $\gamma - \tilde{\gamma}$ using misprediction in the Control group. To solve for $\gamma - \tilde{\gamma}$, we take the expectations over ξ of actual consumption and consumption as predicted at the beginning of the period, difference the two, and rearrange, giving

$$\gamma - \tilde{\gamma} = -\eta m^C. \tag{12}$$

Figure 9 illustrates: naivete $\gamma - \tilde{\gamma}$ is the vertical distance HC between actual and predicted marginal utility, and this can be inferred by multiplying Control group average misprediction m^C (the horizontal distance CL between actual and predicted demand) by the inverse demand slope η.

Temptation

To estimate temptation γ, we take the expectations over ξ of equilibrium consumption in the Limit and Limit Control groups, difference the two, and rearrange, giving

$$\gamma = \eta \tau_2^L. \tag{13}$$

Figure 9 illustrates: temptation γ is the vertical distance LM between desired and actual demand, and this can be inferred by multiplying the effect of removing temptation (τ_2^L, the horizontal distance GL between long-run and present demand) by the inverse demand slope η. We then substitute the estimated γ into equation (12) to infer $\tilde{\gamma}$.

Intercept

Finally, we back out the distribution of ξ that fits the distribution of baseline consumption. We assume that participant i's baseline consumption x_{i1} was in a steady state. Substituting $s_{ss} = \frac{\rho}{1-\rho} x_{ss}$ into equilibrium consumption from equation (6) and rearranging gives

$$\xi_i = p - \gamma + x_{i1}\left(-\eta - \zeta \frac{\rho}{1-\rho}\right). \tag{14}$$

This equation shows that we infer larger ξ_i for people with higher baseline consumption x_{i1}.

In the unrestricted model in Appendix E, equilibrium consumption also depends on ϕ, the direct effect of habit stock on utility. Our data do not allow us to separately identify ϕ from ξ, so we estimate an "intercept" $\kappa_i := (1-\alpha)\delta\rho(\phi - \xi_i) + \xi_i$ that includes both of these structural parameters. In the restricted model with $\alpha = 1$, this simplifies to $\kappa_i = \xi_i$.

6.3 Empirical Moments

Table 3 presents the moments used to estimate the restricted model. The bonus and limit effects τ_t^B and τ_2^L are as displayed in Figure 4. Control group misprediction m^C is the average across periods 2–4 of the difference between actual period t FITSBY use and the prediction for period t elicited on survey t, as displayed in Figure 6. The unrestricted model and our robustness checks also use the anticipatory bonus effect τ_2^B and additional parameters presented in Appendix Table A7. In light of the discussion in Section 5.3, we omit the first half of period 2 when we estimate τ_2^B.[18]

[18] Appendix Table A8 presents parameter estimates when we use all of period 2 to estimate τ_2^B. The estimated projection bias α is smaller, as expected, but the other parameter estimates are very similar.

6.4 Parameter Estimates

Table 4 presents our point estimates and bootstrapped 95 percent confidence intervals. Column 1 presents the restricted model described above (fixing $\tau_2^B = 0$ and $\alpha = 1$), while column 2 presents the unrestricted model described in Appendix E. Since the estimated τ_2^B is close to zero and $\hat{\alpha}$ is close to one, the estimates in the two columns are very similar.

In column 1, we estimate $\hat{\lambda} \approx 1.15$ and $\hat{\rho} \approx 0.299$. In our model, this implies that an exogenous consumption increase of 1 minute per day over a three week period will cause consumption to increase by $\hat{\lambda}\hat{\rho} \approx 0.34$ minutes per day in the next three-week period, and $\hat{\lambda}\hat{\rho}^2 \approx 0.10$ minutes per day in the period after that.

Consistent with the small and statistically insignificant anticipatory bonus effect τ_2^B in the second half of period 2, we estimate $\hat{\alpha} \approx 0.897$ in the unrestricted model in column 2, which is marginally significantly different from one. The point estimate suggests that participants were attentive to only $(1-\hat{\alpha}) \times 100\% \approx$ 10.3 percent of habit formation. Inserting the estimates of λ, ρ, η, and ζ into equation (24) in Appendix E, we calculate that τ_2^B would have needed to be -16.1 minutes per day (compared to the actual point estimate of -1.96 minutes per day in the second half of period 2) to estimate zero projection bias ($\alpha = 0$). In other words, the anticipatory bonus effect is only 12 percent of what our model would predict with fully forward-looking ("rational") habit formation. This is striking when combined with the evidence from Figure 7 that participants correctly predicted habit formation. It is consistent with a model in which people are intellectually aware of habit formation but consume as if they are inattentive to it.

Since the restricted model estimating equations are so simple, one can easily calculate the point estimates in column 1 with the moments from Table 3. For example, the Control group underestimated FITSBY use by an average of 6.13 minutes per day on surveys 2–4. Inserting that into equation (12) gives a naivete of $\widehat{\gamma - \tilde{\gamma}} = -\hat{\eta} \cdot m^C \approx -(-2.68) \cdot (6.13/60) \approx 0.274$ \$/hour in column 1.

The limit changed period 2 FITSBY use by -24.3 minutes per day. Inserting that into equation (13) gives temptation $\hat{\gamma} = \hat{\eta}\tau_2^L \approx (-2.68) \cdot (-24.3/60) \approx 1.09$ \$/hour in column 1. This estimate implies that a tax on FITSBY use of \$1.09 per hour would reduce consumption to the level our participants would choose for themselves in advance. Dividing estimated naivete $\widehat{\gamma - \tilde{\gamma}}$ by this $\hat{\gamma}$ suggests that our participants underestimate temptation by $0.274/1.09 \times 100\% \approx 25$ percent.

Appendix E.5 presents alternative estimates of temptation γ in the restricted and unrestricted models. First, we infer perceived temptation using participants' valuations of the limit functionality and the Screen Time Bonus, following Acland and Levy (2012), Augenblick and Rabin (2019), Chaloupka, Levy, and White (2019), Allcott, Kim, Taubinsky, and Zinman (2021), and Carrera et al. (2021). Second, we generalize the model to include multiple temptation goods, using the self-reports of substitution to FITSBY use on other devices discussed in Section 5.5. Third, we assume that the limit treatment eliminates share $\omega \in [0,1]$ of temptation, relaxing the assumption of $\omega = 1$ in our primary estimates; we estimate ω from differences in self-reported *ideal use change* between the Limit and Limit Control groups. Finally, we allow for individual-specific heterogeneity in γ, using the distribution of *limit tightness* set by Limit group participants. These

alternative approaches all imply temptation γ between about \$1 and \$3 per hour, and our primary estimate of \$1.09 per hour is relatively conservative.

7 Counterfactuals: Effects of Temptation on Time Use

7.1 Methodology

Using the parameter estimates from the previous section, we can predict the effects of changes in temptation and habit formation on steady-state FITSBY use. Equation (21) in Appendix E characterizes steady-state consumption in the unrestricted model. Using that equation, we can predict participant i's steady-state FITSBY use at $p = 0$ as a function of any values of habit formation, temptation, and steady-state misprediction parameters $\{\zeta, \gamma, \tilde{\gamma}, m_{ss}\}$:

$$\hat{x}_{i,ss}(\zeta, \gamma, \tilde{\gamma}, m_{ss}) = \frac{\hat{\kappa}_i + (1-\hat{\alpha})\delta\hat{\rho}\left[(\zeta - \hat{\eta})m_{ss} - \left(1 + \hat{\tilde{\lambda}}\right)\tilde{\gamma}\right] + \gamma}{-\hat{\eta} - (1-\hat{\alpha})\delta\hat{\rho}(\zeta - \hat{\eta}) - \zeta\frac{\hat{\rho} - (1-\hat{\alpha})\delta\hat{\rho}^2}{1-\hat{\rho}}}. \tag{15}$$

The sample average prediction is denoted $\bar{\hat{x}}_{ss}(\zeta, \gamma, \tilde{\gamma}, m_{ss})$. As discussed in Appendix E.3, we assume that the predicted $\tilde{\lambda}$ equals the estimated $\hat{\lambda}$, that steady-state misprediction m_{ss} equals observed Control group misprediction m^C, and that the discount factor is $\delta = 0.997$ per three-week period, consistent with a five percent annual discount rate.

Since we can't identify ϕ (the direct effect of habit stock on utility), we must hold constant each participant's intercept $\kappa_i := (1-\alpha)\delta\rho(\phi - \xi_i) + \xi_i$ across counterfactuals in the restricted model. Since this intercept contains ρ and α, we can't predict consumption with counterfactual values of ρ or α.

In the restricted model with $\alpha = 1$, equation (15) simplifies to

$$\hat{x}_{i,ss}(\rho, \gamma) = \frac{\hat{\xi}_i + \gamma}{-\hat{\eta} - \hat{\zeta}\frac{\rho}{1-\rho}}, \tag{16}$$

which could also be derived from substituting $s_{ss} = \frac{\rho}{1-\rho}x_{ss}$ into equation (6). Steady-state misprediction m_{ss} and perceived temptation $\tilde{\gamma}$ do not affect steady-state consumption in the restricted model because consumers simply maximize current-period flow utility.

7.2 Counterfactual Results

Figure 10 presents point estimates and bootstrapped 95 percent confidence intervals for predicted average FITSBY use at counterfactual parameter values. For each counterfactual, we present predictions from the restricted model ($\alpha = 1$) and unrestricted model ($\alpha = \hat{\alpha}$). We label the restricted model predictions as our primary results, because they are simpler and more conservative.

The first "counterfactual" is the baseline at our point estimates: $\hat{x}_{ss}\left(\hat{\zeta}, \hat{\gamma}, \hat{\tilde{\gamma}}, \hat{m}^C\right)$. This mechanically

matches baseline average FITSBY use of 153 minutes per day. The second counterfactual removes naivete: $\bar{x}_{ss}\left(\hat{\zeta},\hat{\gamma},\hat{\gamma},0\right)$.[19] As described above, naivete has no effect when $\alpha = 1$. Because naivete is so small and projection bias is so strong, the point estimate with $\alpha = \hat{\alpha}$ is very close to the baseline.

The third counterfactual removes temptation: $\bar{x}_{ss}\left(\hat{\zeta},0,0,0\right)$. Relative to baseline, removing temptation reduces predicted FITSBY use by 48 minutes per day (31 percent) with $\alpha = 1$. Thus, our primary estimate is that smartphone FITSBY use would be 31 percent lower without self-control problems.

The fourth and fifth counterfactuals remove habit formation, first with temptation and then without: $\bar{x}_{ss}(0,\hat{\gamma},\hat{\tilde{\gamma}},\hat{m}^C)$ and then $\bar{x}_{ss}(0,0,0,0)$. We emphasize that habit formation on its own is not a departure from rationality (Becker and Murphy 1988), and it could capture forces such as learning and investment that increase consumer welfare. Relative to baseline, removing habit formation reduces predicted FITSBY use by 75 minutes per day with $\alpha = 1$. Without habit formation, the effect of removing temptation (going from the fourth to the fifth counterfactual) is just the limit treatment effect ($\tau_2^L \approx$ -24.3 minutes per day), which is about half of the effect of removing temptation with habit formation (47.5 minutes per day with $\alpha = 1$).[20] This quantifies how habit formation magnifies the effects of temptation, because current temptation increases current consumption and thus future demand.

We highlight one important tension in our results: Figure 4 shows that the limit effects decay slightly over periods 2–5, while our model predicts that the limit effects should grow over time as the Limit group's habit stock diminishes. One potential explanation is that habit formation works differently in response to prices versus self-control tools. Another potential explanation is that motivation to use the limit functionality decays enough that it outweighs the habit stock effect.

Appendix Table A15 presents 19 alternative estimates of the effects of temptation on steady-state FITSBY use across the restricted and unrestricted models. Consistent with the fact that our primary estimates of γ are smaller than most alternative estimates, our primary estimates of the steady-state temptation effects are also relatively conservative. Furthermore, weighting our sample on observables to look more like the U.S. adult population also increases the predicted effects of temptation on consumption. This means that while our sample may still be non-representative on unobservable characteristics, sample selection bias captured by observables causes us to *understate* the effects of temptation on FITSBY use.[21]

Since we don't identify ϕ (the direct effect of habit stock on utility), we can't do a full welfare analysis. The relatively elastic demand—from Section 5.3, 39 percent of consumption is worth less than \$2.50 per hour—suggests that participants do not have strong preferences over how to spend this marginal time, so the welfare losses from self-control problems might be limited. On the other hand, even small individual-level losses might be substantial when aggregated over many social media users. In a static model, the deadweight

[19] Since Figure 7 shows that participants predicted habit formation fairly accurately, we attribute all of steady-state misprediction m_{ss} to naivete about temptation.

[20] Without habit formation, the effect of removing temptation on $\bar{x}_{ss}$ is $\frac{\gamma}{-\eta}$, which equals τ_2^L after substituting $\hat{\gamma} = \tau_2^L \cdot \hat{\eta}$.

[21] Appendix Tables A11–A13 present the demographics, moments, and parameter estimates in the weighted sample. Appendix Table A14 presents the numbers plotted in Figure 10. Appendix Figure A35 presents the distribution of modeled temptation effects across participants, using the Limit group's distribution of *limit tightness* to identify heterogeneity in temptation. The effect is less than 10 minutes per day for 26 percent of participants, and over 100 minutes per day for 13 percent.

loss from temptation would be the triangle *GLM* on Figure 9: $-\tau^L\gamma/2 \approx -(\text{-}24.3/60) \times 1.09/2 \approx \$0.22$ per day, or \$4.62 per three-week period. This is closely consistent with the average valuation of \$4.20 for three weeks of access to the limit functionality. Aggregated across 240 million American social media users (Pew Research Center 2021), this would be $\$4.62 \times (52/3) \times 0.24 \approx \$19.2$ billion per year in welfare losses from overuse of social media caused by self-control problems. For comparison, Facebook's total global profits in 2020 were \$29 billion (United States Securities and Exchange Commission 2020). However, we don't know how these effects would cumulate over time, as represented by ϕ: for example, after a longer period of reduced screen time, people might find more peace of mind or regret the loss of online interactions with friends and family.

8 Conclusion

While digital technologies provide important benefits, some argue that they can be addictive and harmful. We formalize this argument in an economic model and transparently estimate the parameters using data from a field experiment. The Screen Time Bonus intervention had persistent effects after the incentives ended, suggesting that smartphone social media use is habit forming. Participants predicted these persistent effects on surveys but did not reduce FITSBY use before the bonus was in effect, suggesting that they are aware of but inattentive to habit formation. Participants used the screen time limit functionality when we offered it in the experiment, and this functionality reduced FITSBY use by over 20 minutes per day, suggesting that social media use involves self-control problems. The Control group repeatedly underestimated future use, suggesting slight naivete. Many participants reported indicators of smartphone addiction on surveys, and both the bonus and limit interventions reduced this self-reported addiction. Looking at these facts through the lens of our economic model implies that self-control problems magnified by habit formation might be responsible for 31 percent of social media use. These results suggest that better aligning digital technologies with well-being should be an important goal of users, parents, technology workers, investors, and regulators.

Our results raise many additional questions; here are two. First, what are the underlying mechanisms and microfoundations that generate the persistent bonus treatment effects? We model this persistence simply through a capital stock of past consumption, but it could be driven by learning (followed by forgetting), network investments (e.g. connections with friends ebb and flow if maintained or neglected), or more nuanced habit formation mechanisms involving cues or automaticity (e.g. Laibson 2001; Bernheim and Rangel 2004; Steiny Wellsjo 2021). Second, if so many of our participants perceive self-control problems and use (and are willing to pay for) the Phone Dashboard time limit functionality, why isn't there higher demand for commercial digital self-control tools? Only 5 percent of our sample reported using any apps to limit their smartphone use at baseline. Potential explanations include that our experimental setting or selected set of participants overstates demand for commitment, that commercial self-control tools are too expensive or are ineffective because it's too easy to evade them or substitute across devices, that people aren't aware of existing tools, that the time misallocated due to temptation is not very valuable, or that the

commitment and flexibility features we built into Phone Dashboard were better suited to people's needs. We leave these questions for future work.

References

Acland, Dan and Vinci Chow. 2018. "Self-Control and Demand for Commitment in Online Game Playing: Evidence from a Field Experiment." *Journal of the Economic Science Association* 4 (1):46–62.

Acland, Dan and Matthew R. Levy. 2012. "Naivete, Projection Bias, and Habit Formation in Gym Attendance." Working Paper: GSPP13-002.

———. 2015. "Naiveté, Projection Bias, and Habit Formation in Gym Attendance." *Management Science* 61 (1):146–160.

Allcott, Hunt, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow. 2020. "The Welfare Effects of Social Media." *American Economic Review* 110 (3):629–76.

Allcott, Hunt, Matthew Gentzkow, and Lena Song. 2020. "Digital Addiction." AEA RCT Registry. September 27. 10.1257/rct.5796-2.0.

Allcott, Hunt, Joshua Kim, Dmitry Taubinsky, and Jonathan Zinman. 2021. "Are High-Interest Loans Predatory? Theory And Evidence From Payday Lending." Review of Economic Studies, forthcoming.

Allcott, Hunt and Todd Rogers. 2014. "The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation." *American Economic Review* 104 (10):3003–37.

Alter, Adam. 2018. *Irresistible: the Rise of Addictive Technology and the Business of Keeping Us Hooked*. Penguin Press.

Anderson, Michael L. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association* 103 (484):1481–1495.

Andreassen, Cecilie Schou, Torbjørn Torsheim, Geir Scott Brunborg, and Ståle Pallesen. 2012. "Development of a Facebook Addiction Scale." *Psychological Reports* 110 (2):501–517.

Andreoni, James and Charles Sprenger. 2012a. "Estimating Time Preferences from Convex Budgets." *American Economic Review* 102 (7):3333–3356.

———. 2012b. "Risk Preferences Are Not Time Preferences." *American Economic Review* 102 (7):3357–3376.

Ashraf, Nava, Dean Karlan, and Wesley Yin. 2006. "Tying Odysseus to the Mast: Evidence from a Commitment Savings Product in the Philippines." *The Quarterly Journal of Economics* 121 (2):673–697.

Augenblick, Ned. 2018. "Short-Term Discounting of Unpleasant Tasks." Working Paper.

Augenblick, Ned, Muriel Niederle, and Charles Sprenger. 2015. "Working Over Time: Dynamic Inconsistency In Real Effort Tasks." *The Quarterly Journal of Economics* 130 (3):1067–1115.

Augenblick, Ned and Matthew Rabin. 2019. "An Experiment on Time Preference and Misprediction in Unpleasant Tasks." *The Review of Economic Studies* 86 (3):941–975.

Auld, M Christopher and Paul Grootendorst. 2004. "An Empirical Analysis of Milk Addiction." *Journal of Health Economics* 23 (6):1117–1133.

Bai, Liag, Benjamin Handel, Edward Miguel, and Gautam Rao. 2018. "Self-Control and Demand for Preventive Health: Evidence from Hypertension in India." NBER Working Paper No. 23727.

Banerjee, Abhijit and Sendhil Mullainathan. 2010. "The Shape of Temptation: Implications for the Economic Lives of the Poor." NBER Working Paper No. 15973.

Becker, Gary S, Michael Grossman, and Kevin M Murphy. 1994. "An Empirical Analysis of Cigarette Addiction." *The American Economic Review* 84 (3):396–418.

Becker, Gary S and Kevin M Murphy. 1988. "A Theory of Rational Addiction." *Journal of Political Economy* 96 (4):675–700.

Benjamini, Yoav and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society. Series B (Methodological)* 57 (1):289–300.

Bernedo, Maria, Paul J Ferraro, and Michael Price. 2014. "The Persistent Impacts of Norm-Based Messaging and Their Implications for Water Conservation." *Journal of Consumer Policy* 37 (3):437–452.

Bernheim, B. Douglas and Antonio Rangel. 2004. "Addiction and Cue-Triggered Decision Processes." *American Economic Review* 94 (5):1558.

Beshears, John, James J. Choi, Christopher Harris, David Laibson, Brigitte C Madrian, and Jung Sakong. 2015. "Self-Control and Commitment: Can Decreasing the Liquidity of a Savings Account Increase Deposits?" NBER Working Paper No. 21474.

Beshears, John and Katherine Milkman. 2017. "Creating Exercise Habits Using Incentives: The Tradeoff between Flexibility and Routinization." Working Paper, available at https://www.semanticscholar.org/paper/Creating-Exercise-Habits-Using-Incentives-

Bianchi, Adriana and James G Phillips. 2005. "Psychological Predictors of Problem Mobile Phone Use." *CyberPsychology & Behavior* 8 (1):39–51.

Brandon, Alec, Paul J Ferraro, John A List, Robert D Metcalfe, Michael K Price, and Florian Rundhammer. 2017. "Do The Effects of Social Nudges Persist? Theory and Evidence from 38 Natural Field Experiments." NBER Working Paper No. 23277.

Brown, Eileen. 2019. "Americans spend far more time on their smartphones than they think." Available at https://www.zdnet.com/article/americans-spend-far-more-time-on-their-smartphones-than-they-think/.

Bursztyn, Leonardo, Davide Cantoni, David Y Yang, Noam Yuchtman, and Y Jane Zhang. 2020. "Persistent Political Engagement: Social Interactions and the Dynamics of Protest Movements." NBER Conference Paper.

Busse, Meghan R, Devin G Pope, Jaren C Pope, and Jorge Silva-Risso. 2015. "The Psychological Effect of Weather on Car Purchases." *Quarterly Journal of Economics* 130 (1):371–414.

Carrera, Mariana, Heather Royer, Mark Stehr, and Justin Sydnor. 2018. "Can financial incentives help people trying to establish new habits? Experimental evidence with new gym members." *Journal of Health Economics* 58:202–214.

Carrera, Mariana, Heather Royer, Mark Stehr, Justin Sydnor, and Dmitry Taubinsky. 2021. "Who Chooses Commitment? Evidence and Welfare Implications." Working Paper.

Carroll, Gabriel D, James J Choi, David Laibson, Brigitte C Madrian, and Andrew Metrick. 2009. "Optimal Defaults and Active Decisions." *The Quarterly Journal of Economics* 124 (4):1639–1674.

Casaburi, Lorenzo and Rocco Macchiavello. 2019. "Demand and Supply of Infrequent Payments as a Commitment Device: Evidence from Kenya." *American Economic Review* 109 (2):523–555.

Chaloupka, Frank. 1991. "Rational Addictive Behavior and Cigarette Smoking." *Journal of Political Economy* 99 (4):722–742.

Chaloupka, Frank and Kenneth Warner. 1999. "The Economics of Smoking." NBER Working Paper No. 7047.

Chaloupka, Frank J, Matthew R Levy, and Justin S White. 2019. "Estimating Biases in Smoking Cessation: Evidence from a Field Experiment." NBER Working Paper No. 26522.

Charness, Gary and Uri Gneezy. 2009. "Incentives to Exercise." *Econometrica* 77 (3):909–931.

Collis, Avinash and Felix Eggers. 2019. "Effects of Restricting Social Media Usage." Available at SSRN: https://ssrn.com/abstract=3518744.

DellaVigna, Stefano and Ulrike Malmendier. 2006. "Paying Not to Go to the Gym." *American Economic Review* 96 (3):694–719.

Deloitte. 2018. "2018 Global Mobile Consumer Survey: US Edition." Available at https://www2.deloitte.com/content/dam/Deloitte/us/Documents/technology-media-telecommunications/us-tmt-global-mobile-consumer-survey-exec-summary-2018.pdf.

Do, Quy Toan and Hanan G Jacoby. 2020. "Sophisticated Policy with Naive Agents: Habit Formation and Piped Water in Vietnam." Available at SSRN: https://ssrn.com/abstract=3571024.

Duflo, Esther, Michael Kremer, and Jonathan Robinson. 2011. "Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya." *American Economic Review* 101 (6):2350–2390.

Ericson, Keith Marzilli and David Laibson. 2019. *Intertemporal Choice*, vol. 2, chap. 1. Elsevier, 1 ed.

Exley, Christine L. and Jeffrey K. Naecker. 2017. "Observability Increases the Demand for Commitment Devices." *Management Science* 63 (10):3147–3529.

Eyal, Nir. 2020. *Indistractable: How to Control Your Attention and Choose Your Life*. Bloomsbury Publishing PLC.

Fang, Hanming and Dan Silverman. 2004. "Time Inconsistency and Welfare Program Participation: Evidence from the NLSY." Cowles Foundation Discussion Paper No. 1465.

Ferraro, Paul J, Juan Jose Miranda, and Michael K Price. 2011. "The Persistence of Treatment Effects with Norm-Based Policy Instruments: Evidence from a Randomized Environmental Policy Experiment." *American Economic Review* 101 (3):318–22.

Fujiwara, Thomas, Kyle Meng, and Tom Vogl. 2016. "Habit Formation in Voting: Evidence from Rainy Elections." *American Economic Journal: Applied Economics* 8 (4):160–188.

Gerber, Alan S, Donald P Green, and Ron Shachar. 2003. "Voting May be Habit-Forming: Evidence from a Randomized Field Experiment." *American Journal of Political Science* 47 (3):540–550.

Gine, Xavier, Dean Karlan, and Jonathan Zinman. 2010. "Put Your Money Where Your Butt Is: A Commitment Contract for Smoking Cessation." *American Economic Journal: Applied Economics* 2:213–235.

Goda, Gopi Shah, Matthew R. Levy, Colleen Flaherty Manchester, Aaron Sojourner, and Joshua Tasoff. 2015. "The Role of Time Preferences and Exponential-Growth Bias in Retirement Savings." NBER Working Paper No. 21482.

Gosnell, Greer K, John A List, and Robert D Metcalfe. 2020. "The Impact of Management Practices on Employee Productivity: A Field Experiment With Airline Captains." *Journal of Political Economy* 128 (4):1195–1233.

Griffiths, Mark. 2005. "A "Components" Model of Addiction Within a Biopsychosocial Framework." *Journal of Substance Use* 10 (4):191–197.

Gruber, Jonathan and Botond Köszegi. 2001. "Is Addiction "Rational"? Theory and Evidence." *Quarterly Journal of Economics* 116 (4):1261–1303.

Gul, Faruk and Wolfgang Pesendorfer. 2007. "Welfare without Happiness." *American Economic Review* 97 (2):471–476.

Hawley, Josh. 2019. "S. 2314 (116th): SMART Act." Available at https://www.govtrack.us/congress/bills/116/s2314/text.

Hoong, Ruru. 2021. "Self Control and Smartphone Use: An Experimental Study of Soft Commitment Devices." Harvard Working Paper.

Hunt, Melissa G, Rachel Marx, Courtney Lipson, and Jordyn Young. 2018. "No More FOMO: Limiting Social Media Decreases Loneliness and Depression." *Journal of Social and Clinical Psychology* 37 (10):751–768.

Hussam, Reshmaan, Atonu Rabbani, Giovanni Reggiani, and Natalia Rigol. 2019. "Rational Habit Formation: Experimental Evidence from Handwashing in India." Available at SSRN: https://ssrn.com/abstract=3040729.

Irvine, Mark. 2018. "Facebook Ad Benchmarks for YOUR Industry." https://www.wordstream.com/blog/ws/2017/02/28/facebook-advertising-benchmarks.

John, Anett. 2019. "When Commitment Fails - Evidence from a Field Experiment." *Management Science* 66 (2):503–529.

John, Leslie K, George Loewenstein, Andrea B Troxel, Laurie Norton, Jennifer E Fassbender, and Kevin G Volpp. 2011. "Financial Incentives for Extended Weight Loss: a Randomized, Controlled Trial." *Journal of General Internal Medicine* 26 (6):621–626.

Kaur, Supreet, Michael Kremer, and Sendhil Mullainathan. 2015. "Self-Control at Work." *Journal of Political Economy* 123 (6):1227–1277.

Kemp, Simon. 2020. "Digital 2020 Reports." Available at wearesocial.com/digital-2020.

Kuchler, Theresa and Michaela Pagel. 2018. "Sticking to Your Plan: The Role of Present Bias for Credit Card Paydown." NBER Working Paper No. 24881.

Laibson, David. 1997. "Golden Eggs and Hyperbolic Discounting." *Quarterly Journal of Economics* 112 (2):443–478.

———. 2001. "A Cue-Theory of Consumption." *The Quarterly Journal of Economics* 116 (1):81–119.

———. 2018. "Private Paternalism, the Commitment Puzzle, and Model-Free Equilibrum." *AEA Papers and Proceedings* 108:1–21.

Laibson, David, Peter Maxted, Andrea Repetto, and Jeremy Tobacman. 2015. "Estimating Discount Functions with Consumption Choices over the Lifecycle." Working Paper.

Levitt, Steven D, John A List, and Sally Sadoff. 2016. "The Effect of Performance-Based Incentives on Educational Achievement: Evidence from a Randomized Experiment." NBER Working Paper No. 22107.

Liu, Zhuang, Michael Sockin, and Wei Xiong. 2020. "Data Privacy and Temptation." NBER Working Paper No. 27653.

Loewenstein, George, Ted O'Donoghue, and Matthew Rabin. 2003. "Projection Bias in Predicting Future Utility." *Quarterly Journal of Economics* 118 (4):1209–1248.

Madrian, Brigitte C and Dennis F Shea. 2001. "The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior." *The Quarterly Journal of Economics* 116 (4):1149–1187.

Makarov, Uliana. 2011. "Networking or not working: A model of social procrastination from communication." *Journal of Economic Behavior & Organization* 2011 (80):574–585.

Marotta, Veronica and Alessandro Acquisti. 2017. "Online Distractions, Website Blockers, and Economic Productivity: A Randomized Field Experiment." Preliminary Draft.

Mosquera, Roberto, Mofioluwasademi Odunowo, Trent McNamara, Xiongfei Guo, and Ragan Petrie. 2019. "The Economic Effects of Facebook." *Experimental Economics* :1–28.

New York Post. 2017. "Americans Check Their Phones 80 Times a Day: Study." Available at https://nypost.com/2017/11/08/americans-check-their-phones-80-times-a-day-study/.

Newport, Cal. 2019. *Digital Minimalism: Choosing a Focused Life in a Noisy World*. Penguin Random House.

O'Donoghue, Ted and Matthew Rabin. 1999. "Doing It Now or Later." *American Economic Review* 89 (1):103–124. URL https://www.aeaweb.org/articles?id=10.1257/aer.89.1.103.

Paserman, M. Daniele. 2008. "Job Search and Hyperbolic Discounting: Structural Estimation and Policy Evaluation." *The Economic Journal* 118:1418–1452.

Pew Research Center. 2021. "Social Media Fact Sheet." URL https://www.pewresearch.org/internet/fact-sheet/social-media/.

Read, Danieal and Barbara Van Leeuwen. 1998. "Predicting Hunger: The Effects of Appetite and Delay on Choice." *Organizational Behavior and Human Decision Processes* 76 (2):189–205.

Rees-Jones, Alex and Kyle T. Rozema. 2020. "Price Isn't Everything: Behavioral Response around Changes in Sin Taxes."

Royer, Heather, Mark Stehr, and Justin Sydnor. 2015. "Incentives, Commitments, and Habit Formation in Exercise: Evidence from a Field Experiment with Workers at a Fortune-500 Company." *American Economic Journal: Applied Economics* 7 (3):51–84.

Sadoff, Sally, Anya Savikhin Samek, and Charles Sprenger. 2020. "Dynamic Inconsistency in Food Choice: Experimental Evidence from a Food Desert." *Review of Economic Studies* 87 (4):1954–1988.

Sagioglu, Christina and Tobias Greitemeyer. 2014. "Facebook's Emotional Consequences: Why Facebook Causes a Decrease in Mood and Why People Still Use It." *Computers in Human Behavior* 35:359–363.

Schilbach, Frank. 2019. "Alcohol and Self-Control: A Field Experiment in India." *American Economic Review* 109 (4):1290–1322.

Shapiro, Jesse M. 2005. "Is There a Daily Discount Rate? Evidence from the Food Stamp Nutrition Cycle." *Journal of Public Economics* 89:303–325.

Shui, Haiyan and Lawrence M Ausubel. 2005. "Time Inconsistency in the Credit Card Market." Working Paper.

Skiba, Paige Marta and Jeremy Tobacman. 2018. "Payday Loans, Uncertainty, and Discounting: Explaining Patterns of Borrowing, Repayment, and Default." Working Paper.

Steiny Wellsjo, Alexandra. 2021. "Simple Actions, Complex Habits: Lessons from Hospital Hand Hygiene." Access on https://drive.google.com/file/d/1wbn6IuU0tMQ2VN6YHSWSCXv4v9pucKyK/view.

Strack, Philipp and Dmitry Taubinsky. 2021. "Dynamic Preference "Reversals" and Time Inconsistency." NBER Working Paper No. 28961.

Toussaert, Severine. 2018. "Eliciting Temptation and Self-Control Through Menu Choices: A Lab Experiment." *Econometrica* 86 (3):859–889.

Tromholt, Morten. 2016. "The Facebook Experiment: Quitting Facebook Leads to Higher Levels of Well-Being." *Cyberpsychology, Behavior, and Social Networking* 19 (11):661–666.

United States Securities and Exchange Commission. 2020. "Official Facebook 2020 10-K report as filed with SEC." Available at https://d18rn0p25nwr6d.cloudfront.net/CIK-0001326801/4dd7fa7f-1a51-4ed9-b9df-7f42cc3321eb.pdf.

Van Soest, Daan and Ben Vollaard. 2019. "Breaking Habits." Working Paper.

Vanman, Eric J, Rosemary Baker, and Stephanie J Tobin. 2018. "The Burden of Online Friends: The Effects of Giving Up Facebook on Stress and Well-Being." *The Journal of Social Psychology* 158 (4):496–507.

Vox. 2020. "Tech Companies Tried to Help us Spend Less Time on Our Phones. It Didn't Work." Available at https://www.vox.com/recode/2020/1/6/21048116/tech-companies-time-well-spent-mobile-phone-usage-data.

World Health Organization. 2018. "Gaming Disorder." Available at https://www.who.int/features/qa/gaming-disorder/en/.

Wurmser, Yoram. 2020. "US Mobile Time Spent 2020." Available at https://www.emarketer.com/content/us-mobile-time-spent-2020.

Zenith Media. 2019. "Consumers Will Spend 800 Hours Using Mobile Internet Devices This Year." Available at https://www.zenithmedia.com/consumers-will-spend-800-hours-using-mobile-internet-devices-this-year/.

Table 1: **Experiment Timeline and Sample Sizes**

Phase	Date	Sample size
Recruitment and intake	March 22 - April 8	3,271,165 shown ads 26,101 clicked on ads 18,589 passed screen 8,514 consented 5,320 finished intake survey
Survey 1 (baseline)	April 12	4,134 began Survey 1 4,038 finished Survey 1 2,126 were randomized
Survey 2	May 3	2,068 began Survey 2 2,053 informed of treatment, of which: 2,048 were not in MPL group 2,032 finished Survey 2
Survey 3	May 24	1,993 began Survey 3 1,981 finished Survey 3
Survey 4	June 14	1,954 began Survey 4 1,948 finished Survey 4
Completion	July 26	1,938 kept Phone Dashboard through July 26, of which: 1,933 were not in MPL group ("analysis sample")

Table 2: **Sample Demographics**

	(1) Analysis sample	(2) U.S. adults
Income ($000s)	40.8	43.0
College	0.67	0.30
Male	0.39	0.49
White	0.72	0.74
Age	33.7	47.6
Period 1 phone use (minutes/day)	333.0	.
Period 1 FITSBY use (minutes/day)	152.8	.

Notes: Column 1 presents average demographics for our analysis sample, and column 2 presents average demographics of American adults using data from the 2018 American Community Survey.

Table 3: **Empirical Moments for Restricted Model Estimation**

Parameter	Description	(1) Point estimate	(2) Confidence interval
τ_3^B	Contemporaneous bonus effect (minutes/day)	-55.9	[-61.7, -50.3]
τ_4^B	Long-term bonus effect (minutes/day)	-19.2	[-24.7, -13.7]
τ_5^B	Long-term bonus effect (minutes/day)	-12.3	[-18.1, -6.54]
τ_2^L	Limit effect (minutes/day)	-24.3	[-28.1, -20.4]
m^C	Control group misprediction (minutes/day)	6.13	[4.52, 7.72]
$\bar{x}_1$	Average baseline use (minutes/day)	153	[149, 157]

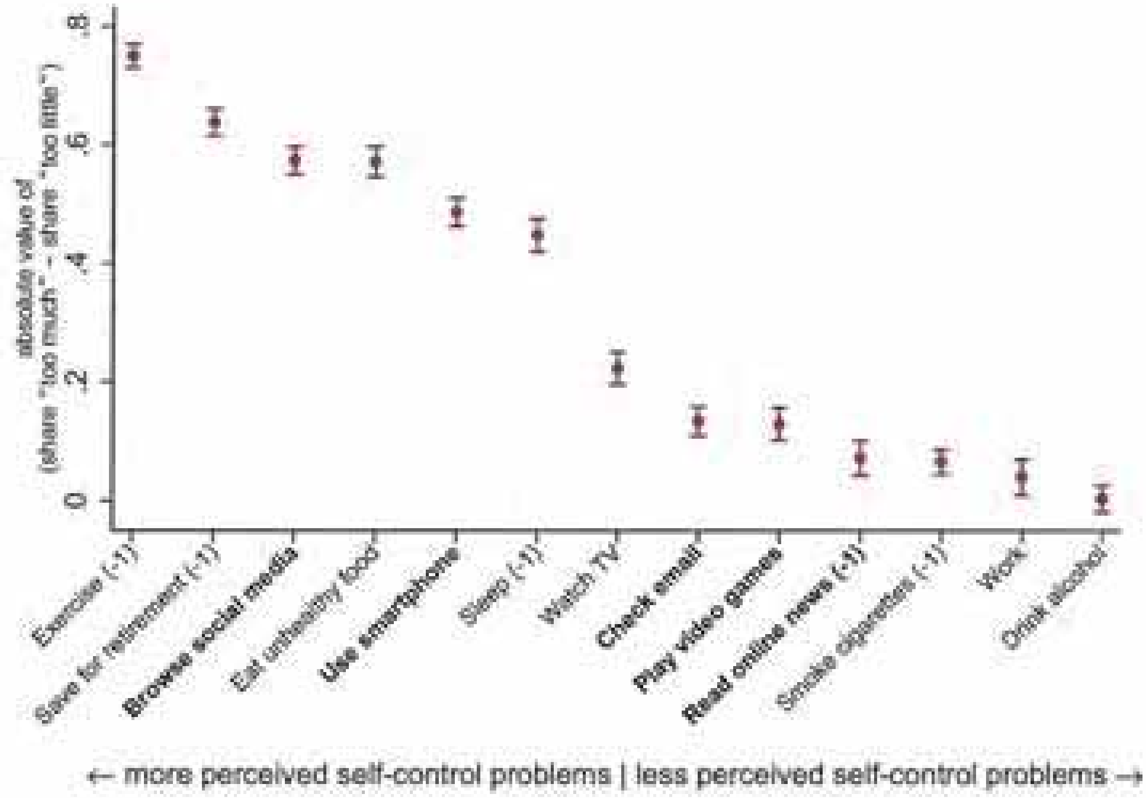
Notes: This table presents point estimates and bootstrapped 95 percent confidence intervals for the empirical moments used for our primary estimates of the restricted model.

Table 4: **Primary Parameter Estimates**

Parameter	Description (units)	(1) Restricted model ($\tau_2^B = 0, \alpha = 1$)	(2) Unrestricted model ($\alpha = \hat{\alpha}$)
λ	Habit stock effect on consumption (unitless)	1.15	1.12
		$[0.609, 3.31]$	$[0.572, 3.16]$
ρ	Habit formation (unitless)	0.299	0.302
		$[0.106, 0.493]$	$[0.106, 0.498]$
α	Projection bias (unitless)	1	0.897
			$[0.584, 1.00]$
η	Price coefficient ($-day/hour^2)	-2.68	-2.75
		$[-2.98, -2.43]$	$[-3.04, -2.51]$
ζ	Habit stock effect on marginal utility ($-day/hour^2)	3.08	3.01
		$[1.65, 8.97]$	$[1.55, 8.57]$
$\gamma - \tilde{\gamma}$	Naivete about temptation ($/hour)	0.274	0.278
		$[0.201, 0.349]$	$[0.205, 0.354]$
γ	Temptation ($/hour)	1.09	1.11
		$[0.884, 1.30]$	$[0.903, 1.33]$
$\bar{\kappa}$	Average intercept ($/hour)	-2.41	-2.24
		$[-3.62, -1.10]$	$[-3.53, -0.803]$

Notes: This table presents point estimates and bootstrapped 95 percent confidence intervals from the estimation strategy described in Section 6.2 and Appendix E.3.

Figure 1: **Online and Offline Temptation**



Notes: This figure presents responses to the following question, which we asked participants in our experiment during the baseline survey, "For each of the activities below, please tell us whether you think you do it too little, too much, or the right amount." The bars are ordered from left to right in order of largest to smallest absolute value of (share "too little" – share "too much").

Figure 2: **Experimental Design**

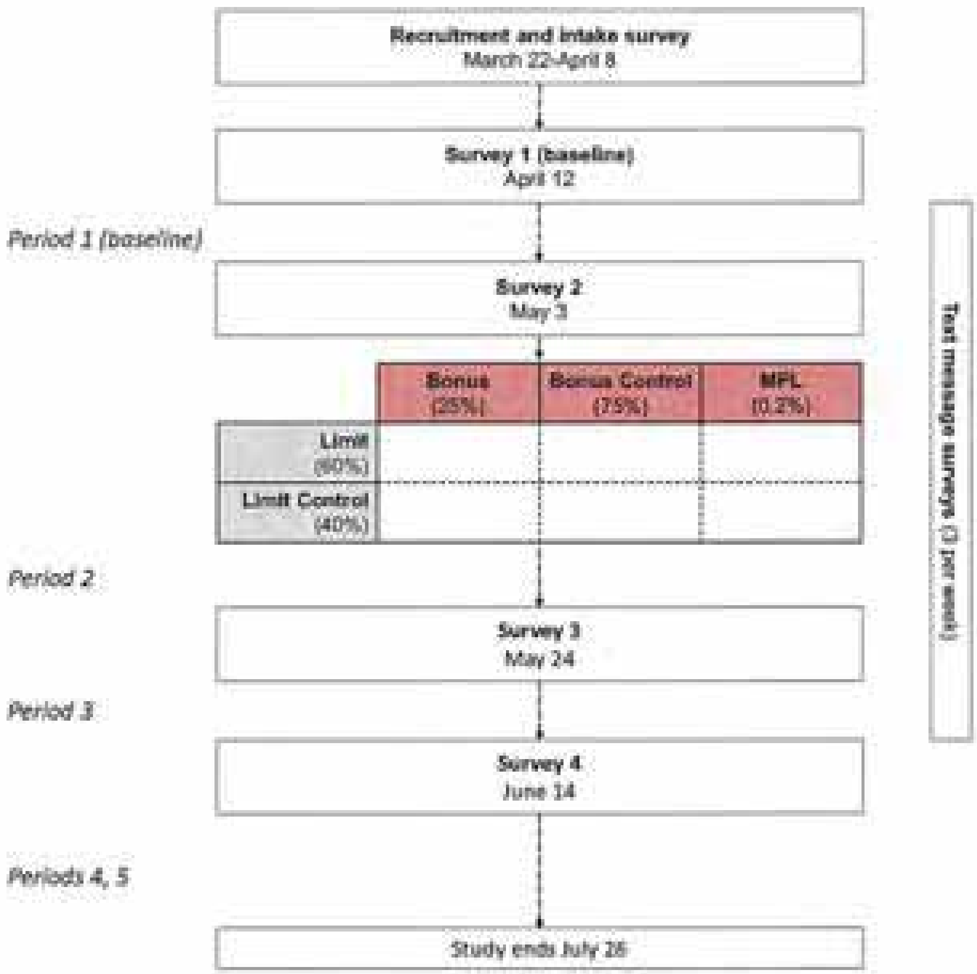
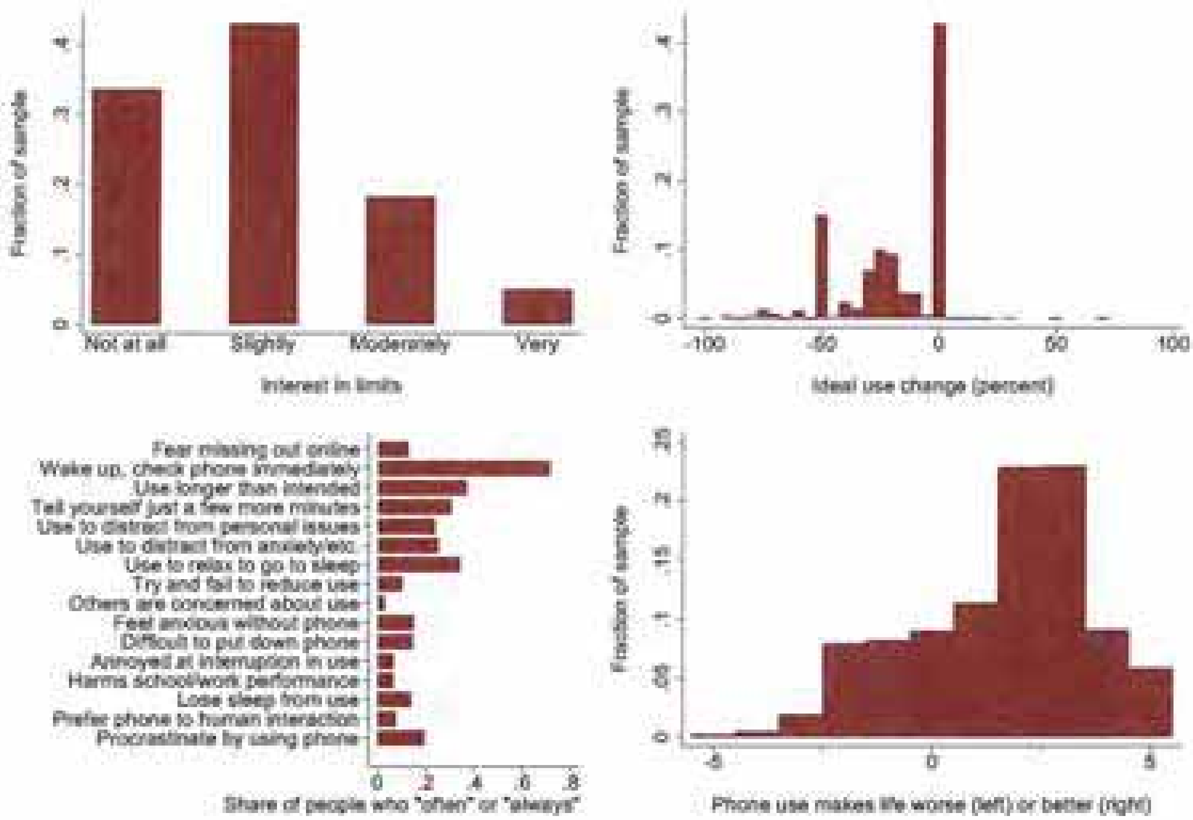
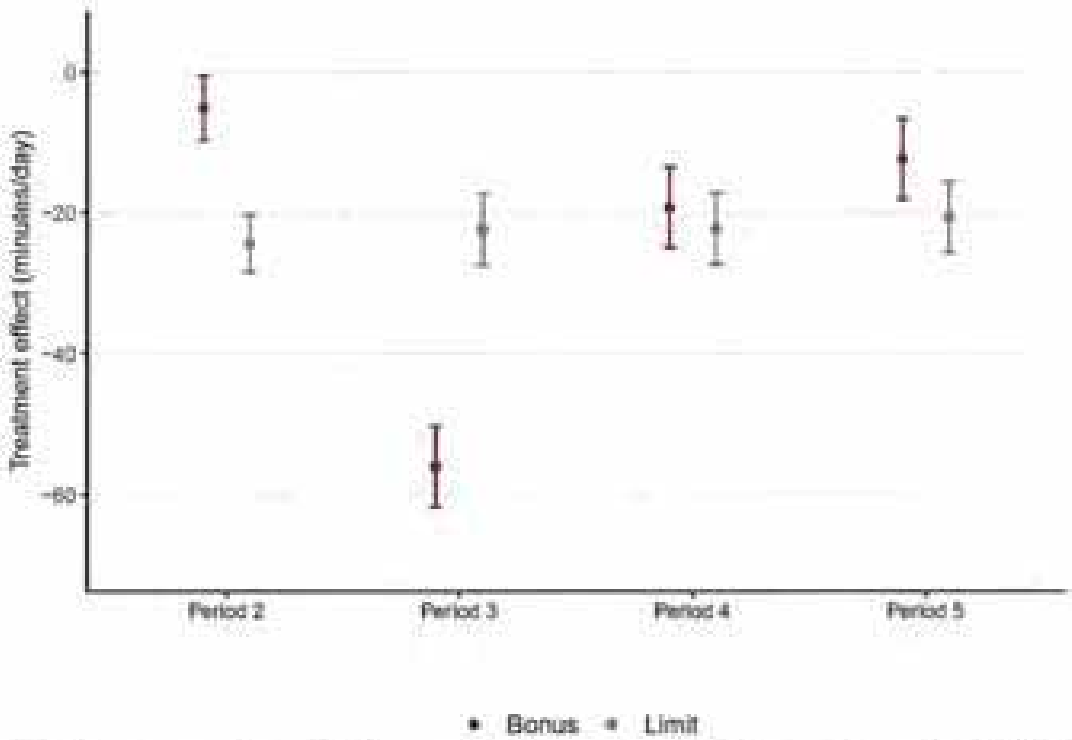


Figure 3: **Baseline Qualitative Evidence of Self-Control Problems**



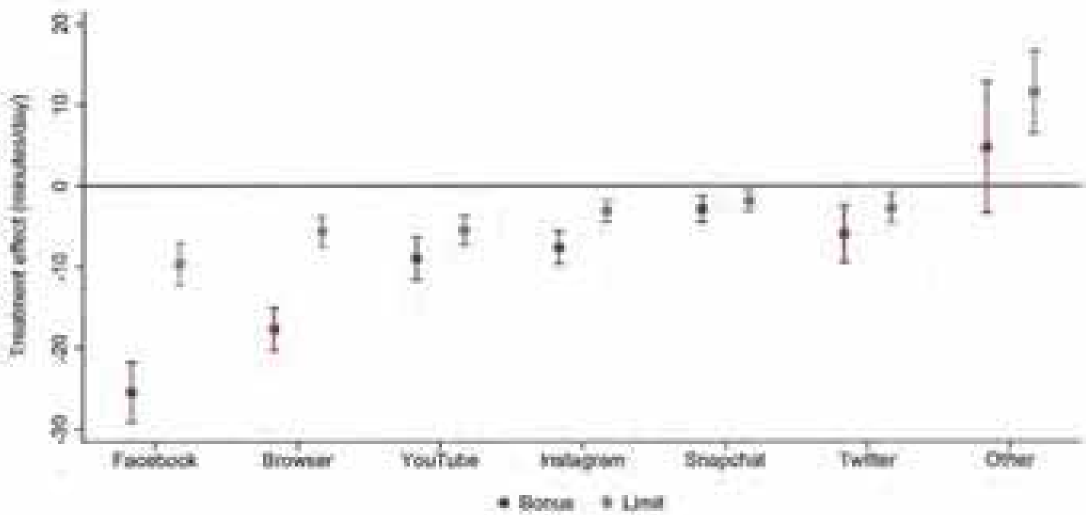
Notes: This figure presents the distributions of four measures of smartphone addiction from the baseline survey. *Interest in limits* is the answer to, "How interested are you to set limits on your phone use?" *Ideal use change* is the answer to, "Relative to your actual use over the past 3 weeks, by how much would you ideally have [reduced/increased] your screen time?" The bottom left panel presents the share of participants who responded "often" or "always" to each of 16 questions modified from the Mobile Phone Problem Use Scale and the Bergen Facebook Addiction Scale. *Phone use makes life worse or better* is the answer to, "To what extent do you think your smartphone use made your life better or worse over the past 3 weeks?"

Figure 4: **Treatment Effects on FITSBY Use**



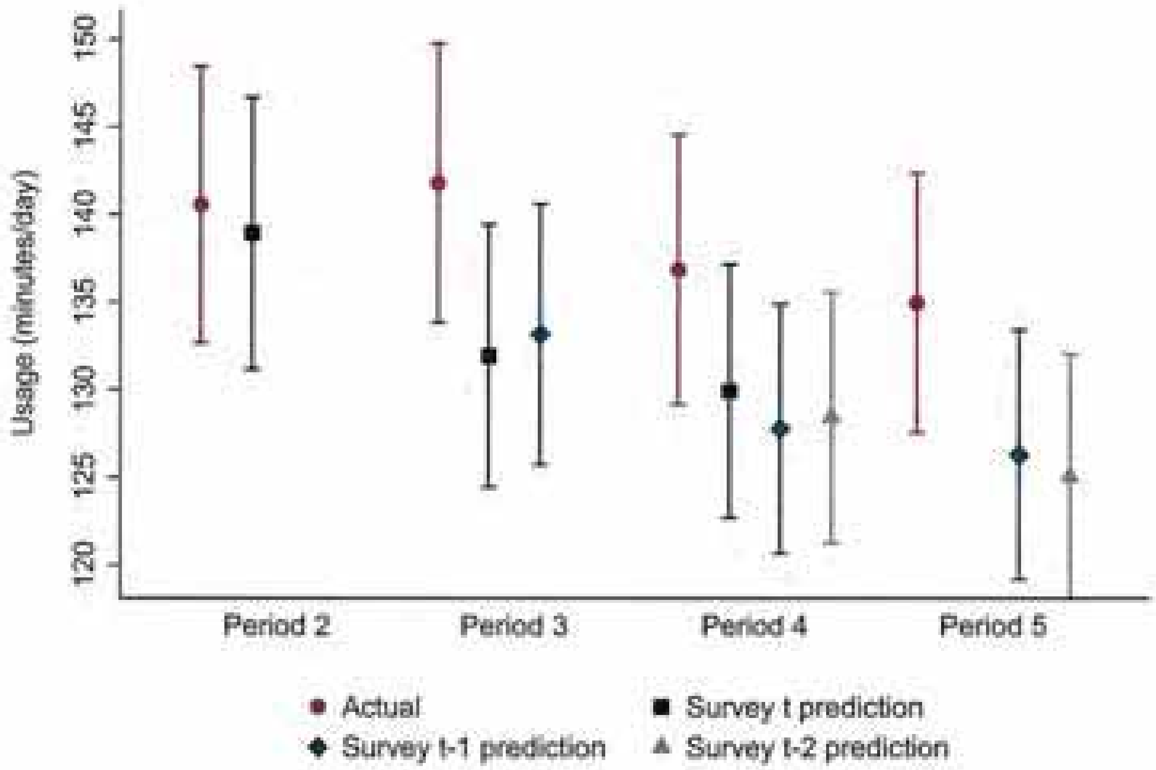
Notes: This figure presents effects of the bonus and limit treatments on FITSBY use using equation (4). FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers, and YouTube.

Figure 5: **Effects on Smartphone Use by App**



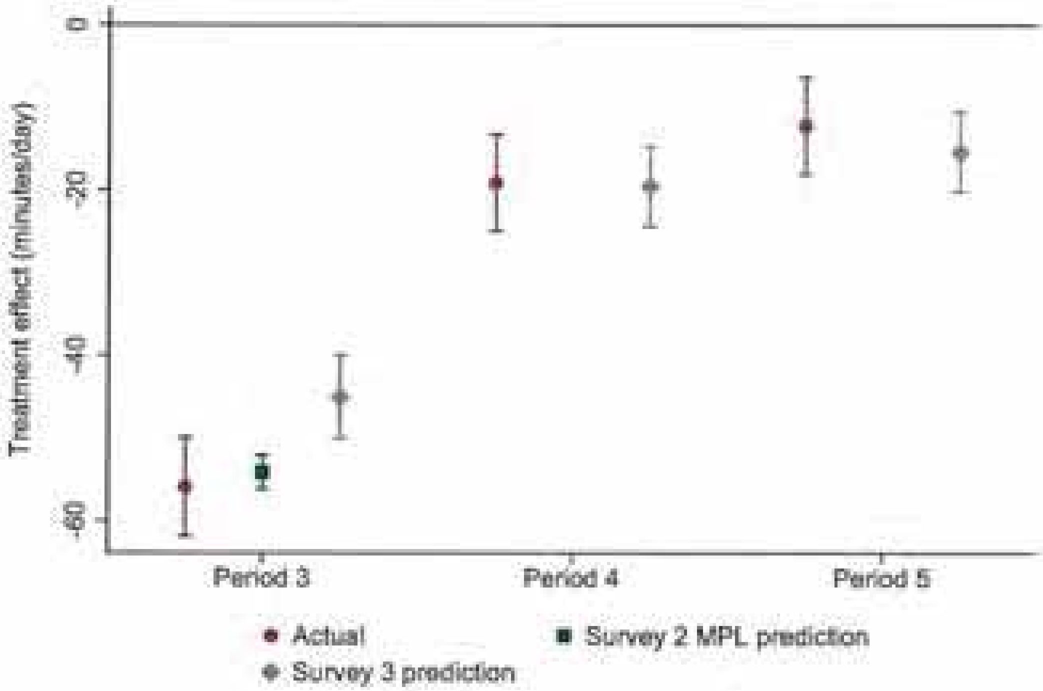
Notes: This figure presents effects of the bonus and limit treatments on smartphone use by app using equation (4). The bonus effects are measured in period 3, while the limit effects are measured in periods 2–5. FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers, and YouTube. FITSBY apps are in order of decreasing period 1 use.

Figure 6: **Predicted vs. Actual FITSBY Use in Control Conditions**



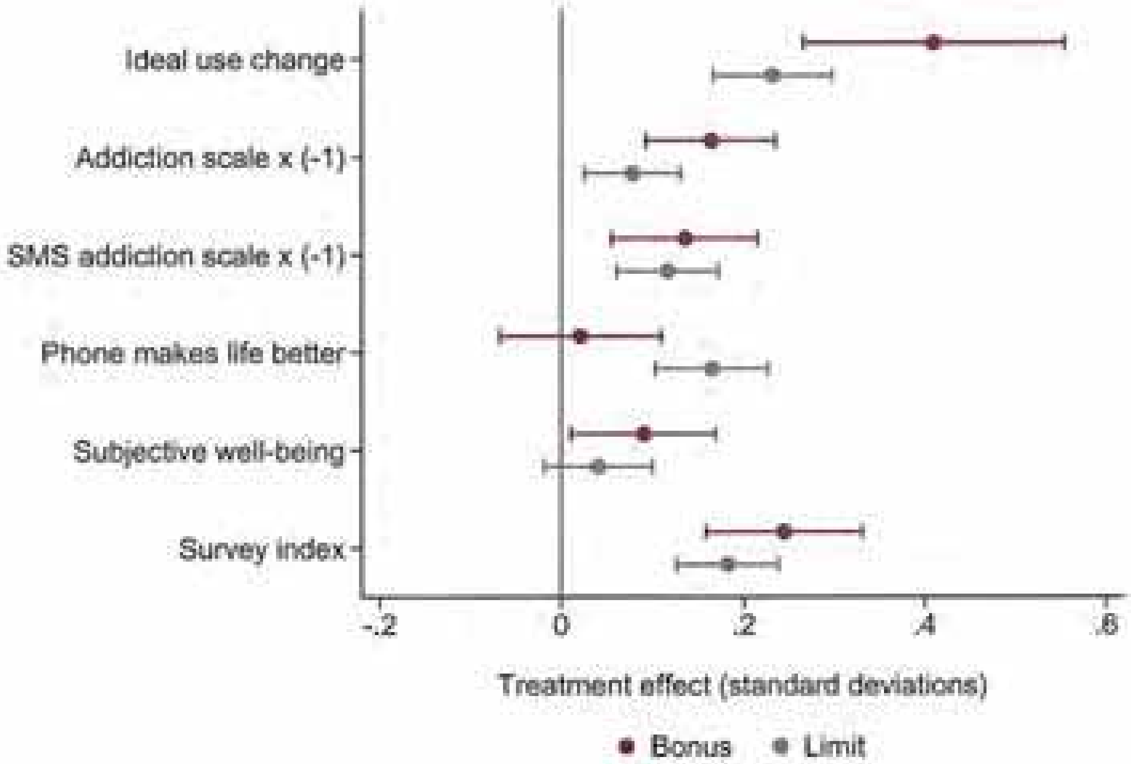
Notes: This figure presents average actual FITSBY use by period and average predicted FITSBY use for that period, for participants in the intersection of the Bonus Control and Limit Control groups. Period t is the three weeks immediately after survey t, so "survey t prediction" is the prediction for period t made just prior to period t. FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers, and YouTube.

Figure 7: **Predicted vs. Actual Habit Formation**



Notes: This figure presents the treatment effects of the bonus on FITSBY use and on predicted FITSBY use from survey 3 using equation (4), as well as the average predicted bonus treatment effect elicited on survey 2 before the bonus multiple price list. FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers, and YouTube.

Figure 8: **Effects of Limits and Bonus on Survey Outcome Variables**



Notes: This figure presents effects of the bonus and limit treatments on survey outcome variables using equation (4). The bonus effect is measured on survey 4, while the limit effect is measured on both surveys 3 and 4. *Ideal use change* is the answer to, "Relative to your actual use over the past 3 weeks, by how much would you ideally have [reduced/increased] your screen time?" *Addiction scale* is answers to a battery of 16 questions modified from the Mobile Phone Problem Use Scale and the Bergen Facebook Addiction Scale. *SMS addiction scale* is answers to shortened versions of the addiction scale questions delivered via text message. *Phone makes life better* is the answer to, "To what extent do you think your smartphone use made your life better or worse over the past 3 weeks?" *Subjective well-being* is answers to seven questions reflecting happiness, life satisfaction, anxiety, depression, concentration, distraction, and sleep quality; anxiety, depression, and distraction are re-oriented so that more positive reflects better subjective well-being. *Survey index* combines the previous five variables, weighting by the inverse of their covariance at baseline.

Figure 9: **Model Identification**

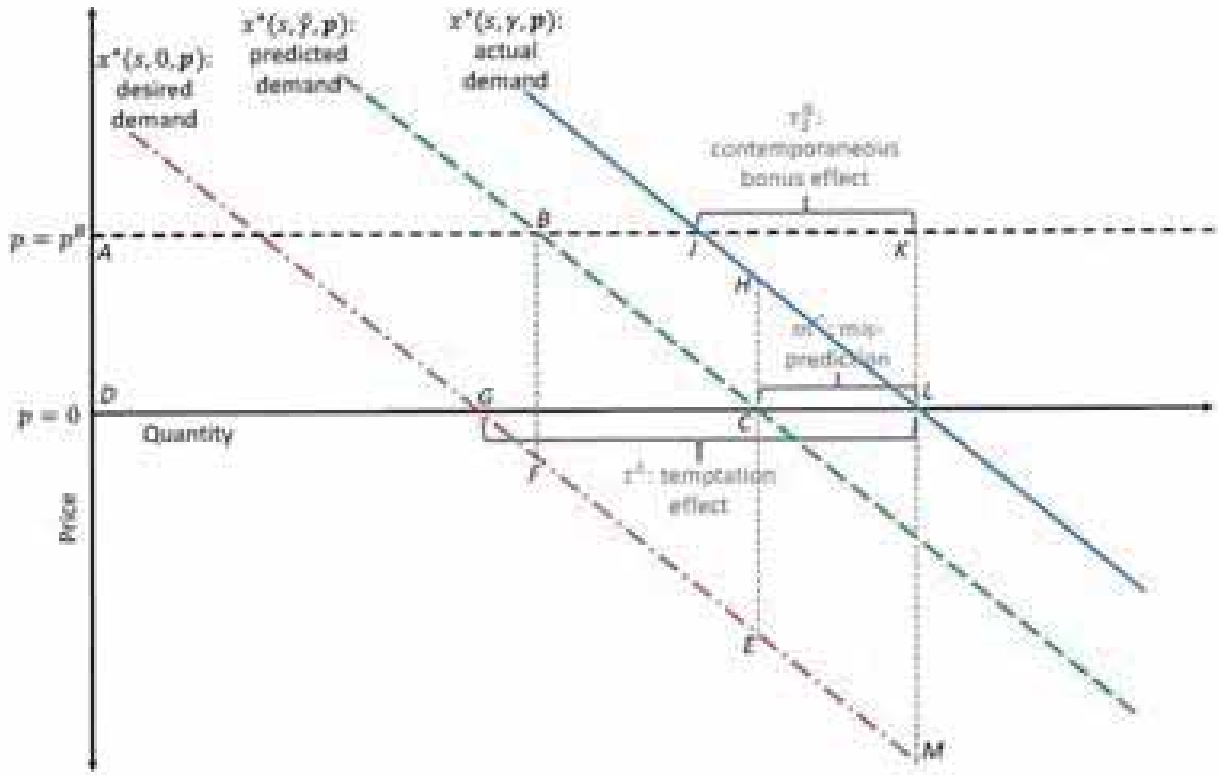
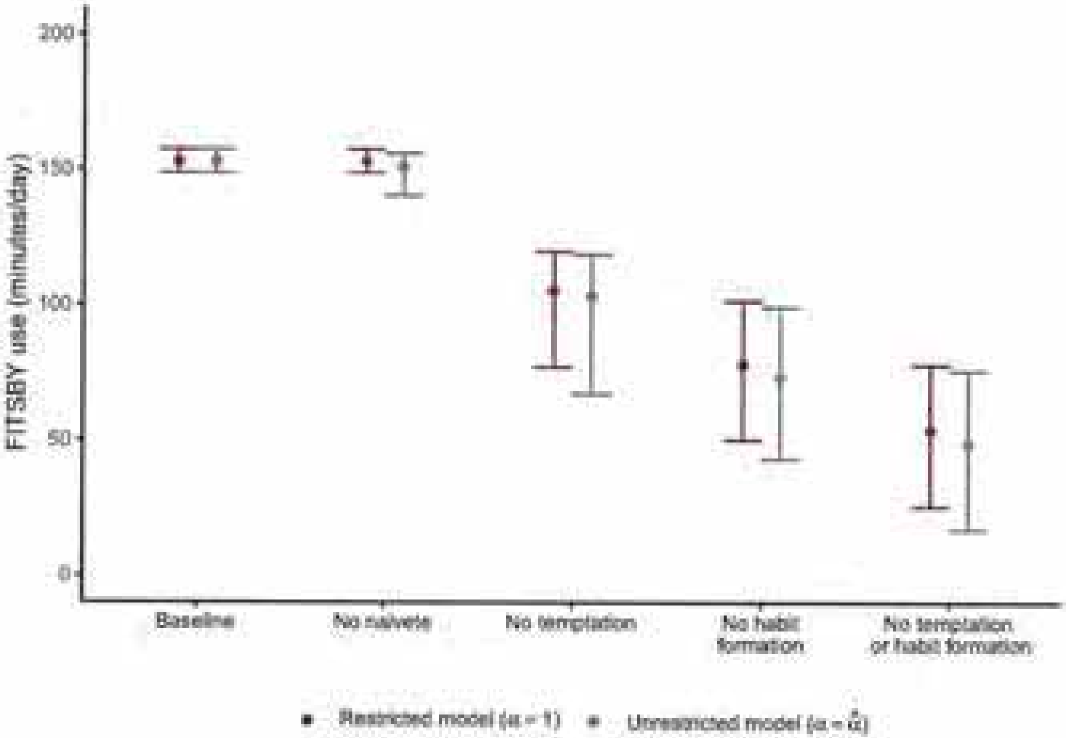


Figure 10: **Effects of Temptation and Habit Formation on FITSBY Use**



Notes: This figure presents point estimates and bootstrapped 95 percent confidence intervals for predicted steady-state FITSBY use with different parameter assumptions, using equation (15).

Online Appendix

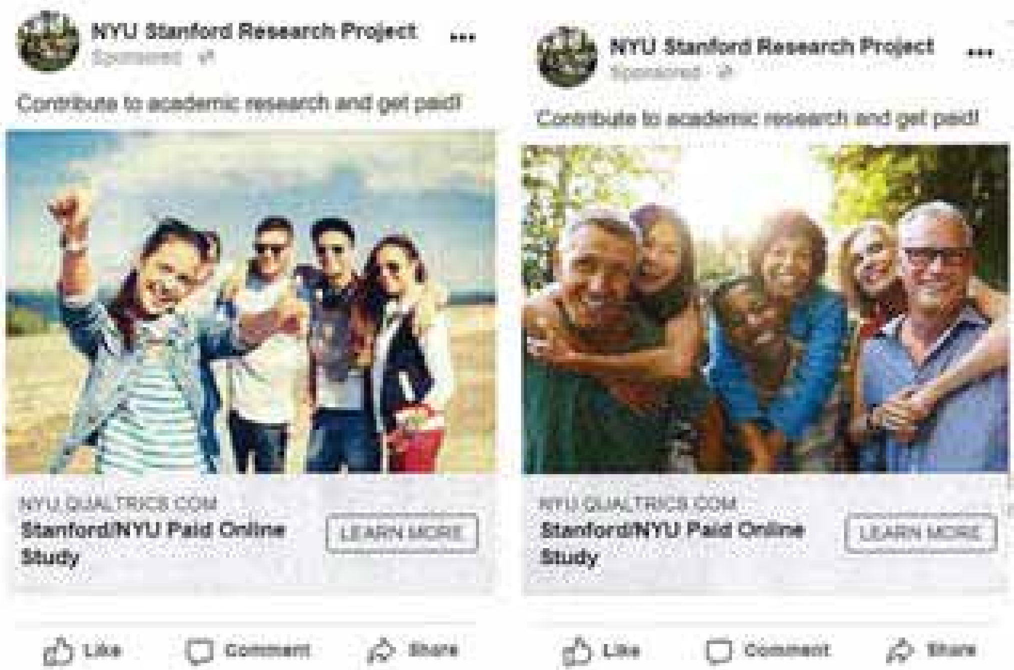
Digital Addiction

Hunt Allcott, Matthew Gentzkow, and Lena Song

Table of Contents

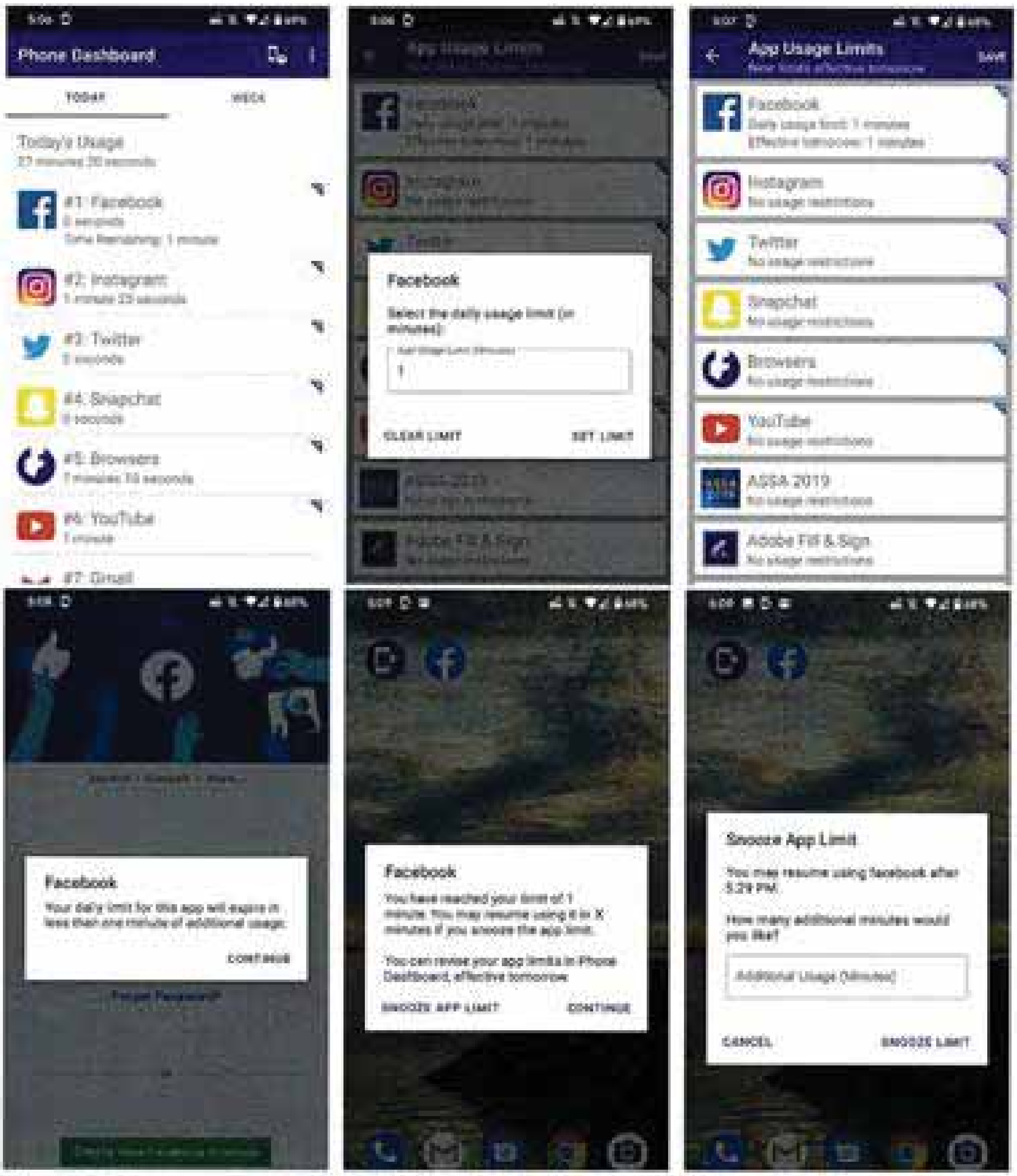
A Experimental Design Appendix

Figure A1: **Facebook Recruitment Ads**



Notes: The ads at left and right were shown to users aged 18–34 and 35–64, respectively.

Figure A2: **Phone Dashboard Screenshots**



Notes: This figure presents screenshots of the Phone Dashboard app. The top left presents the day's total usage by app. The top middle shows how a user can set daily a daily usage limit for each app, effective tomorrow. The top right shows the usage limits set for each app. The bottom left shows the warning users receive when they are within five minutes or one minute of their limit. The bottom middle shows the message users receive when they reach the limit. Users with the snooze functionality can resume using an app after a delay of $X \in \{0, 2, 5, 20\}$ minutes. The bottom right shows the option for a user to choose how many additional minutes to add to the daily limit after the snooze delay. All participants had the usage information in the top left panel, while only the Limit group had the time limit functionalities in the other panels.

A.1 Variable Definitions

Ideal use change. Some people say they use their smartphone too much and ideally would use it less. Other people are happy with their usage or would ideally use it more. How do you feel about your overall smartphone use over the past 3 weeks?

- I used my smartphone too much.
- I used my smartphone the right amount.
- I used my smartphone too little.

Relative to your actual use over the past 3 weeks, by how much would you ideally have [if "too much": reduced. If "too little": increased] your smartphone use? Please give a number in percent. ___ %

Addiction scale. Over the past 3 weeks, how often have you...

- Been worried about missing out on things online when not checking your phone?
- Checked social media, text messages, or email immediately after waking up?
- Used your phone longer than intended?
- Found yourself saying "just a few more minutes" when using your phone?
- Used your phone to distract yourself from personal problems?
- Used your phone to distract yourself from feelings of guilt, anxiety, helplessness, or depression?
- Used your phone to relax in order to go to sleep?
- Tried to reduce your phone use without success?
- Experienced that people close to you are concerned about the amount of time you use your phone?
- Felt anxious when you don't have your phone?
- Found it difficult to switch off or put down your phone?
- Been annoyed or bothered when people interrupt you while you use your phone?
- Felt your performance in school or at work suffers because of the amount of time you use your phone?
- Lost sleep due to using your phone late at night?
- Preferred to use your phone rather than interacting with your partner, friends, or family?
- Put off things you have to do by using your phone?

Never, Rarely, Sometimes, Often, Always

SMS addiction scale.

- In the past 24 hours, did you use your phone longer than intended?
- In the past 24 hours, did your performance at school or work suffer because of the amount of time you used your phone?
- In the past 24 hours, did you feel like you had an easy time controlling your screen time?
- In the past 24 hours, did you use your phone mindlessly?
- In the past 24 hours, did you use your phone because you were feeling down?
- In the past 24 hours, did using your phone keep you from working on something you needed to do?
- In the past 24 hours, would you ideally have used your phone less?
- Last night, did you lose sleep because of using your phone late at night?
- When you woke up today, did you immediately check social media, text messages, or email?

Please text back your answer on a scale from 1 (not at all) to 10 (definitely).

Phone makes life better. To what extent do you think your smartphone use made your life better or worse over the past 3 weeks?

11-point scale from -5 (Makes my life worse) to 0 (Neutral) to 5 (Makes my life better)

Subjective well-being. Please tell us the extent to which you agree or disagree with each of the following statements. Over the past 3 weeks, ...

- ... I was a happy person
- ... I was satisfied with my life
- ... I felt anxious
- ... I felt depressed
- ... I could concentrate on what I was doing
- ... I was easily distracted
- ... I slept well

7-point scale from strongly disagree to neutral to strongly agree

B Data Appendix

Table A1: **Response Rates**

(a) **Limit**

	(1) Control	(2) All limits	(3) Snooze 0	(4) Snooze 2	(5) Snooze 5	(6) Snooze 20	(7) No snooze	(8) F-test p-value
Completed survey 3	0.97	0.96	0.96	0.98	0.96	0.97	0.95	0.51
Completed survey 4	0.95	0.94	0.95	0.95	0.94	0.95	0.93	0.81
Have period 2 usage	1.00	1.00	0.99	0.99	1.00	1.00	1.00	0.23
Have period 3 usage	0.99	0.98	0.99	0.98	0.99	0.98	0.98	0.68
Have period 4 usage	0.98	0.97	0.99	0.98	0.97	0.97	0.96	0.37
Have period 5 usage	0.97	0.96	0.97	0.96	0.96	0.97	0.95	0.70

(b) **Bonus**

	(1) Control	(2) Treatment	(3) t-test p-value
Completed survey 3	0.97	0.96	0.74
Completed survey 4	0.95	0.95	0.64
Have period 2 usage	1.00	1.00	0.16
Have period 3 usage	0.98	0.98	0.95
Have period 4 usage	0.98	0.97	0.84
Have period 5 usage	0.96	0.96	0.85

Notes: Columns 1 and 2 of Panel (a) present present response rates for Limit and Limit Control groups. Columns 3–7 present response rates for each of the snooze delay conditions within the Limit group. Column 8 presents the p-value of an F-test of differences between the Limit Control and the separate snooze delay conditions. Columns 1 and 2 of Panel (b) present response rates for Bonus and Bonus Control groups. Column 3 presents the p-value of a t-test of differences between the Bonus and Bonus Control groups.

Table A2: **Covariate Balance**

(a) **Limit**

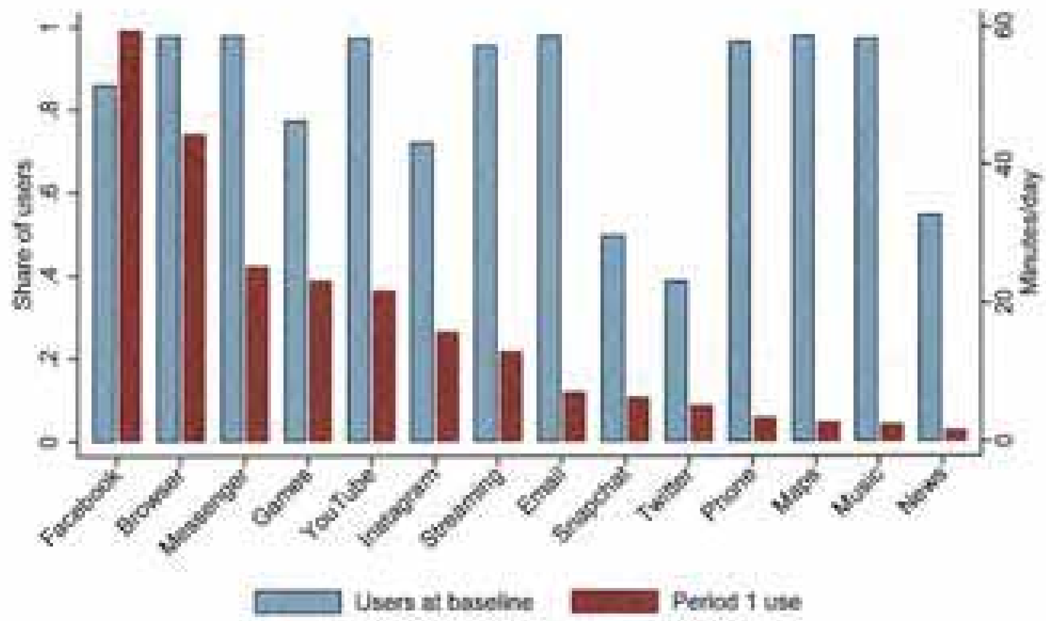
Variable	(1) Treatment Mean/SD	(2) Control Mean/SD	t-test p-value (1)-(2)
Income ($000s)	40.15 (36.22)	41.76 (37.84)	0.35
College	0.67 (0.47)	0.67 (0.47)	0.72
Male	0.38 (0.49)	0.40 (0.49)	0.51
White	0.70 (0.46)	0.74 (0.44)	0.13
Age	33.61 (12.33)	33.79 (12.35)	0.76
Period 1 FITSBY use (minutes/day)	151.96 (92.00)	154.07 (99.19)	0.64
N	1150	783	
F-test of joint significance (p-value)			0.65
F-test, number of observations			1933

(b) **Bonus**

Variable	(1) Treatment Mean/SD	(2) Control Mean/SD	t-test p-value (1)-(2)
Income ($000s)	41.26 (39.16)	40.65 (36.11)	0.76
College	0.67 (0.47)	0.67 (0.47)	0.75
Male	0.41 (0.49)	0.38 (0.49)	0.26
White	0.71 (0.46)	0.72 (0.45)	0.61
Age	33.53 (12.17)	33.73 (12.40)	0.76
Period 1 FITSBY use (minutes/day)	151.24 (91.97)	153.34 (95.94)	0.67
N	479	1454	
F-test of joint significance (p-value)			0.94
F-test, number of observations			1933

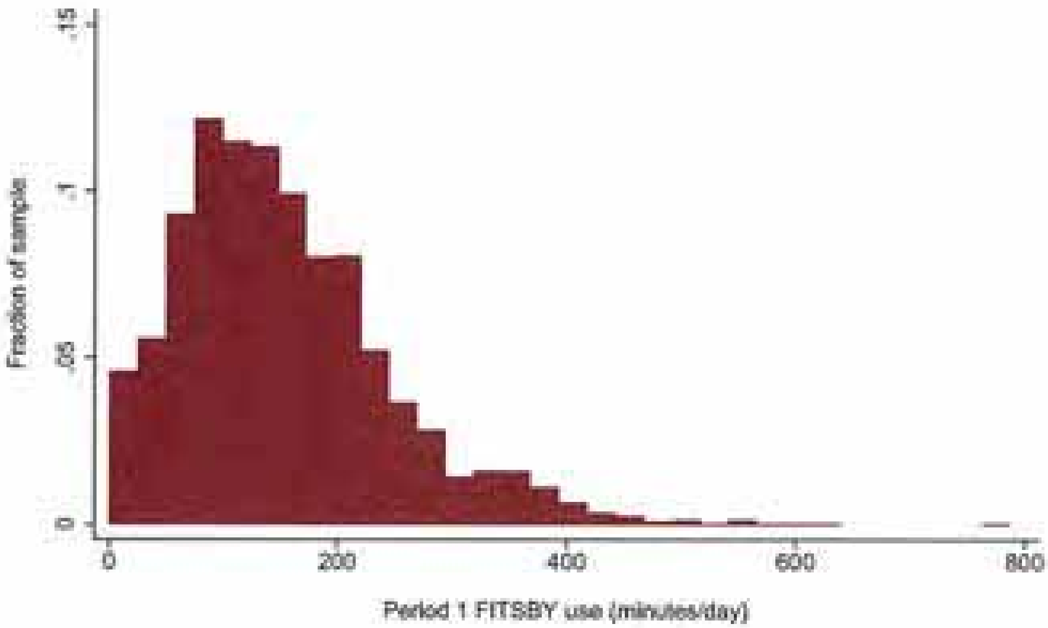
Notes: Panels (a) and (b) present tests of covariate balance for the Limit and Bonus treatment and control groups.

Figure A3: **Most Popular Apps**



Notes: This figure presents the share of users that have each app and the average daily screen time in period 1 (baseline). Period 1 use is across all users, not conditioning on whether or not they have the app.

Figure A4: **Distribution of Baseline FITSBY Use**



Notes: This figure presents a distribution of FITSBY use in period 1 (baseline). FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers, and YouTube.

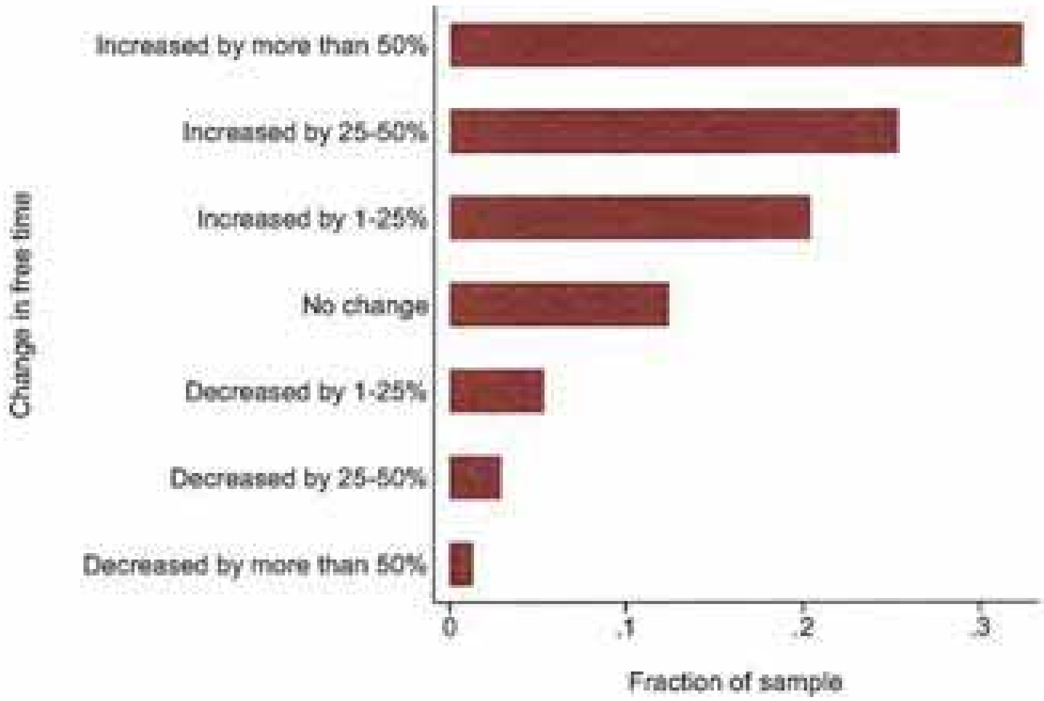
Table A3: **Descriptive Statistics for Survey Outcome Variables**

	Mean	Standard deviation	Minimum value	Maximum value
Ideal use change	-19.0	21.4	-100	70
Addiction scale x (-1)	-6.2	2.6	-16	0
SMS addiction scale x (-1)	1.7	3.1	-9	9
Phone makes life better	1.6	2.0	-5	5
Subjective well-being	0.2	2.5	-7	7

Notes: This table present descriptive statistics for the survey outcome variables at baseline.

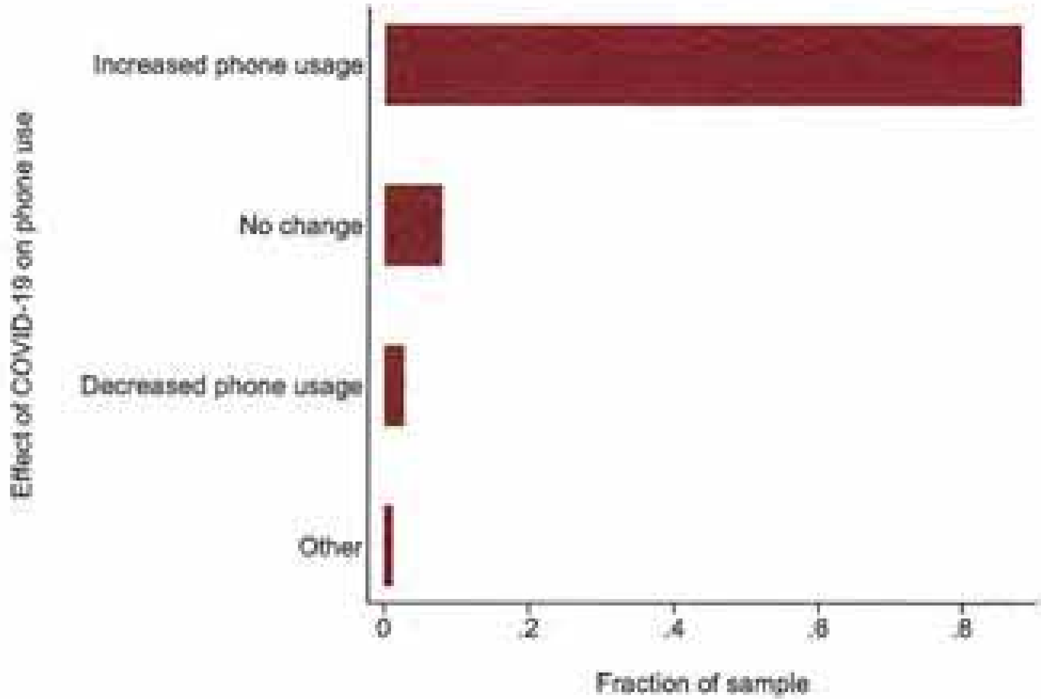
C Differences Between 2019 and the Study Period

Figure A5: **Effects of Coronavirus Outbreak on Free Time**



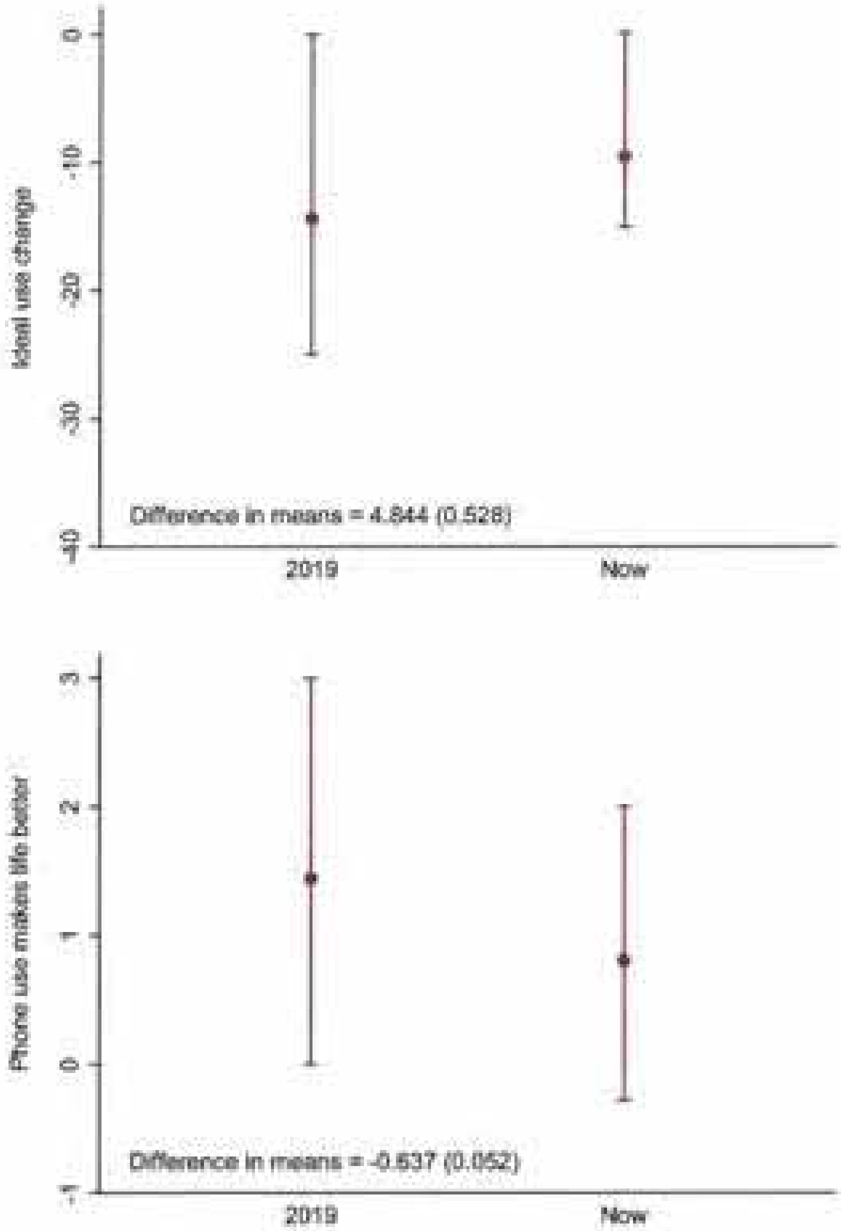
Notes: This figure presents the distribution of responses to the baseline survey question, "To what extent has the recent coronavirus outbreak changed how much free time you have?"

Figure A6: **Effects of Coronavirus on Smartphone Use**



Notes: The baseline survey asked, "How has the recent coronavirus outbreak changed how you use your smartphone?" We coded the responses as to whether they indicated increased, decreased, or unchanged smartphone use.

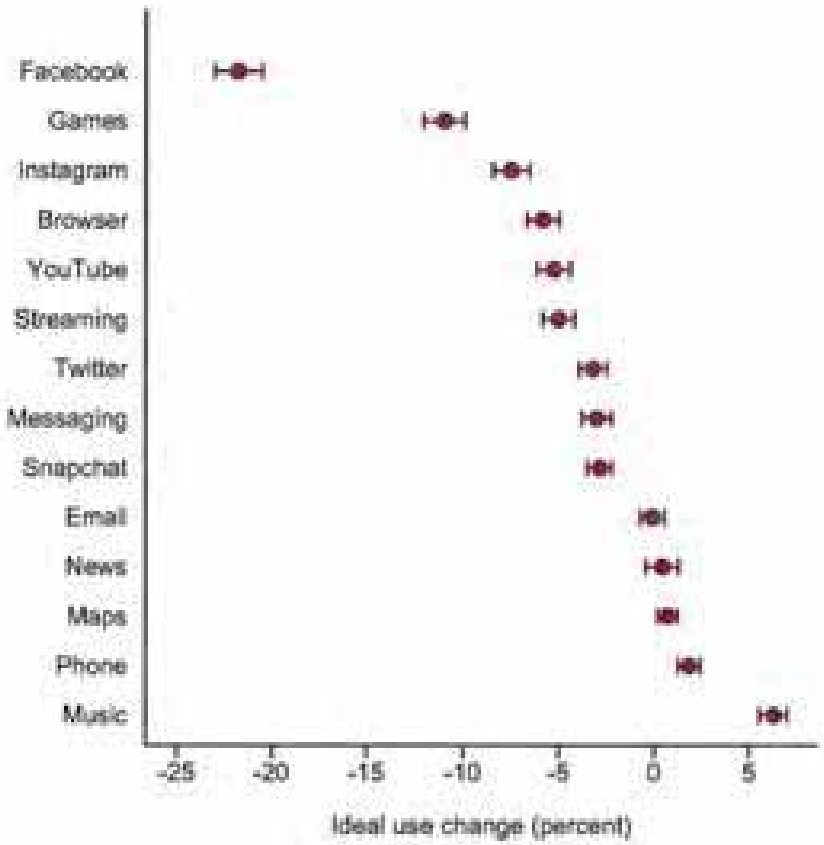
Figure A7: **Self-Control Problems in 2019 versus Now**



Notes: This figure presents the mean (dots) and 25th and 75th percentiles (spikes) of responses to *ideal use change* and *phone use makes life better* for 2019 and for the past 3 weeks, as reported on the baseline survey. *Ideal use change* is the answer to, "Relative to your actual use [in 2019 / over the past 3 weeks], by how much would you ideally have [reduced/increased] your screen time? *Phone use makes life better* is the answer to, "To what extent do you think your smartphone use made your life better or worse [in 2019 / over the past 3 weeks]?"

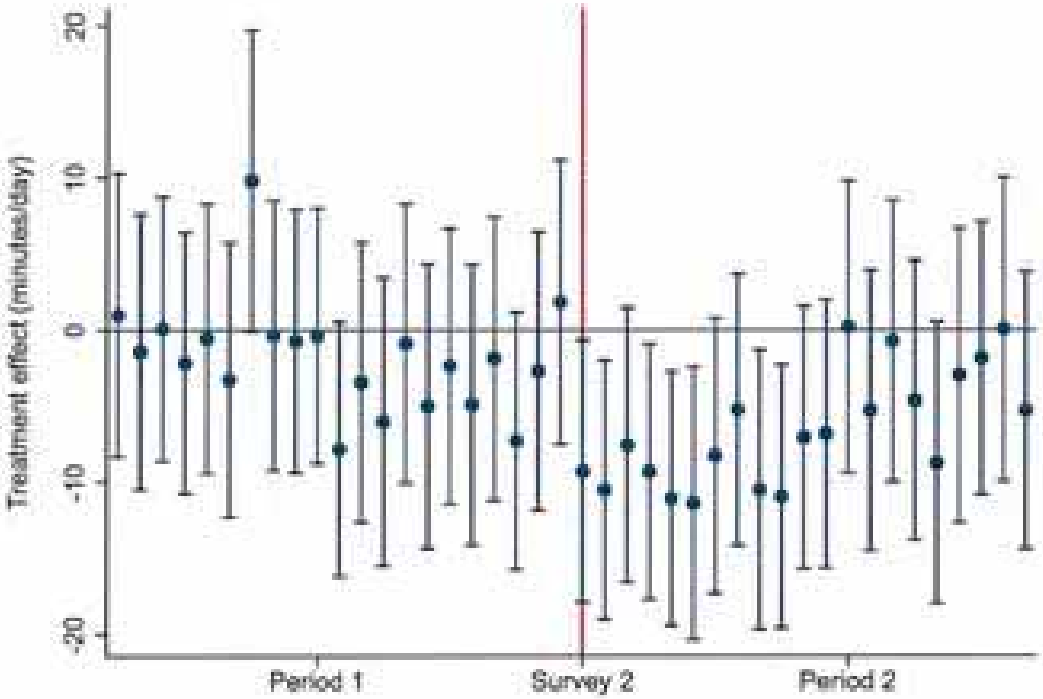
D Model-Free Results Appendix

Figure A8: **Ideal Use Change by App or Category**



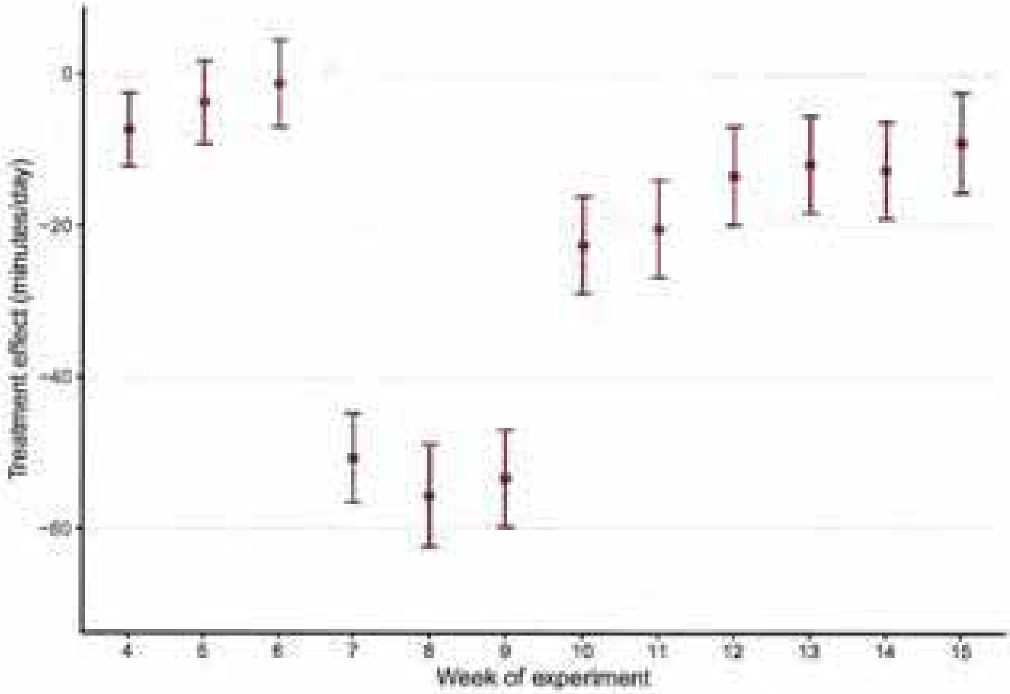
Notes: This figure presents mean *ideal use change* by app or app category at baseline. *Ideal use change* is the answer to, "Relative to your actual use over the past 3 weeks, by how much would you ideally have [reduced/increased] your screen time?" We code "I don't use this app at all" as 0, so these results reflect how much each app contributes to overall temptation, not how tempting each app is for the subset of people who use it.

Figure A9: **Effects of Bonus on FITSBY Use by Day for Periods 1 and 2**



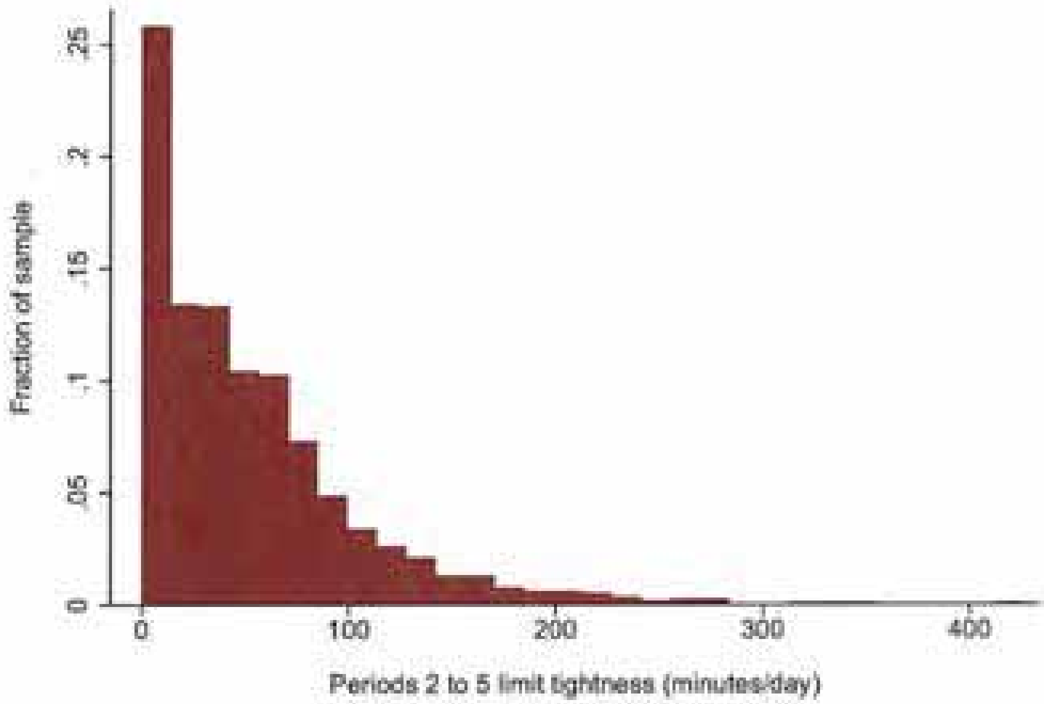
Notes: This figure presents differences in average FITSBY use between the Bonus and Bonus Control group for each day of periods 1 and 2. The vertical line indicates the day of survey 2, when the bonus was announced. FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers, and YouTube.

Figure A10: **Effects of Bonus on FITSBY Use by Week**



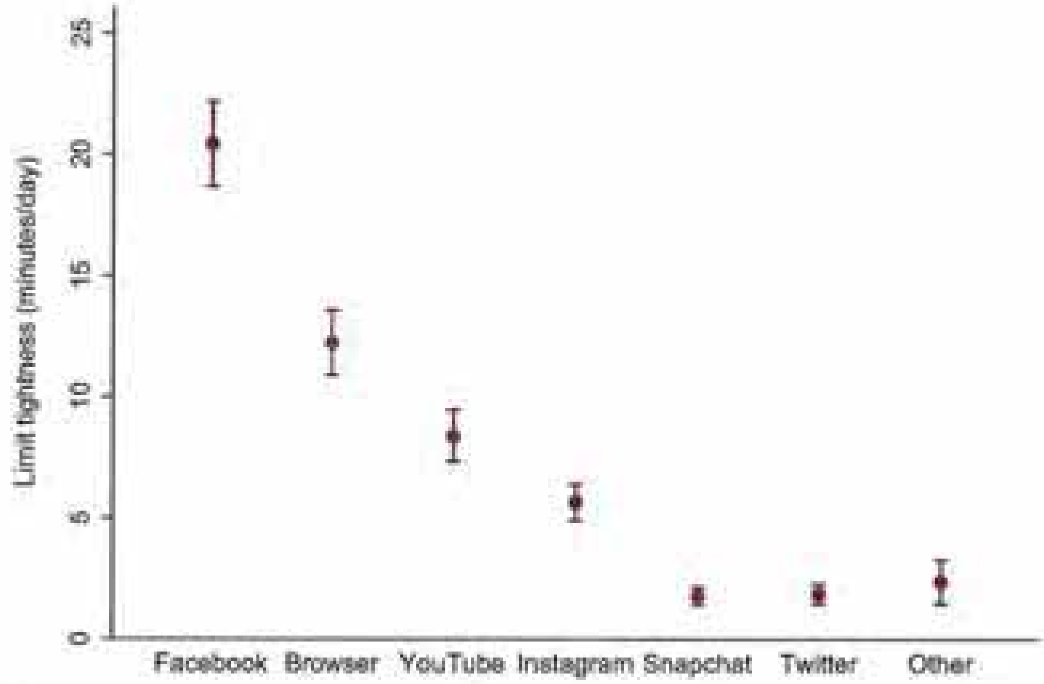
Notes: This figure presents effects of the bonus treatment on FITBSY use by week using equation (4). FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers, and YouTube.

Figure A11: **Distribution of User-Level Limit Tightness**



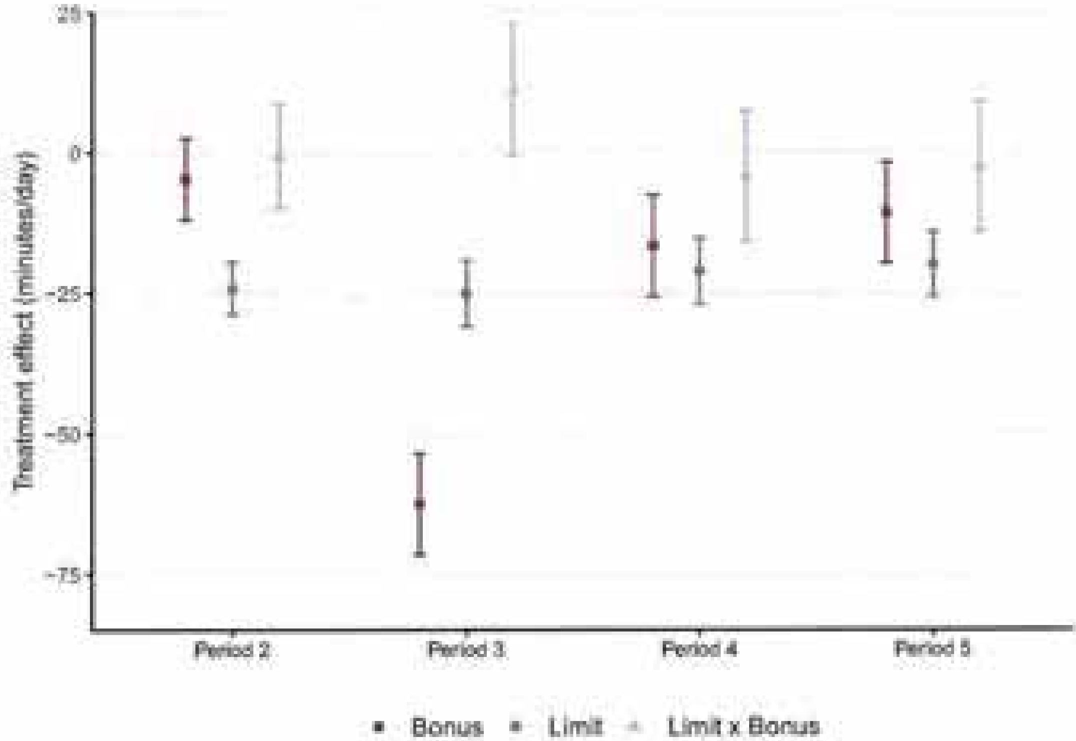
Notes: This figure presents mean *user-level limit tightness* over periods 2–5. *User-level limit tightness* is the amount by which a user's limits would have hypothetically reduced overall screen time if applied to their baseline use without snoozes; see equation (5).

Figure A12: **Average Limit Tightness by App**



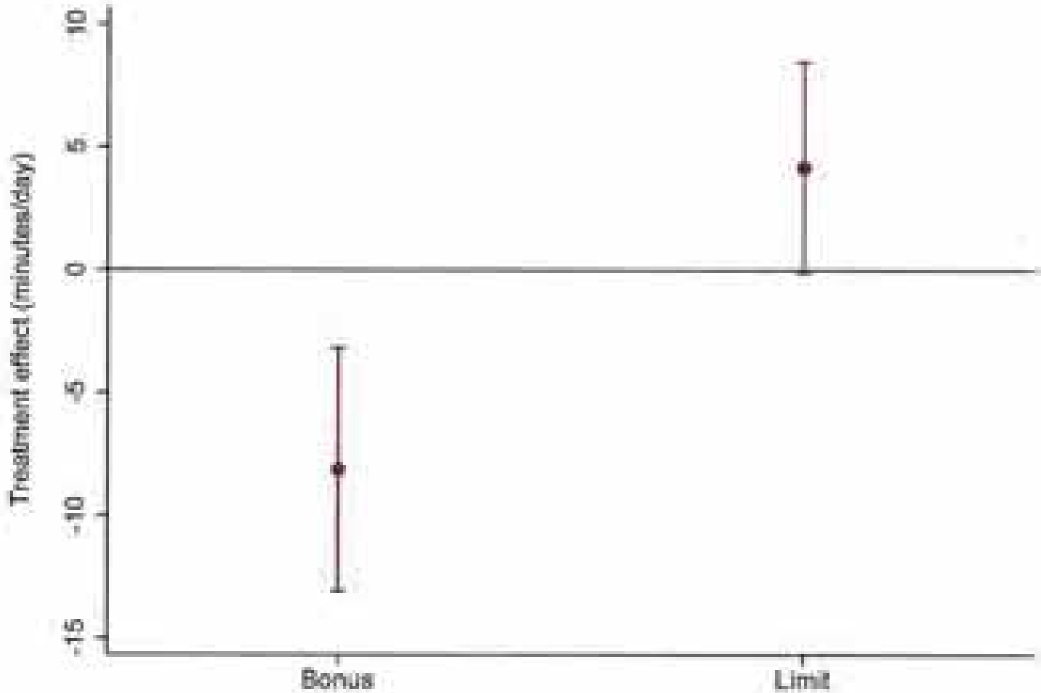
Notes: This figure presents average *limit tightness* by app over periods 2–5. *Limit tightness* is the amount by which a user's limits would have hypothetically reduced screen time if applied to their baseline use without snoozes; see equation (5). FITSBY apps are in order of decreasing period 1 use.

Figure A13: **Interaction Effects of Bonus and Limit by Period**



Notes: This figure presents effects of bonus and limit treatments on FITSBY use using equation (4) with an additional interaction term for participants in the intersection of the Limit and Bonus groups. FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers, and YouTube.

Figure A14: **Effects on Self-Reported FITSBY Use Change on Other Devices**



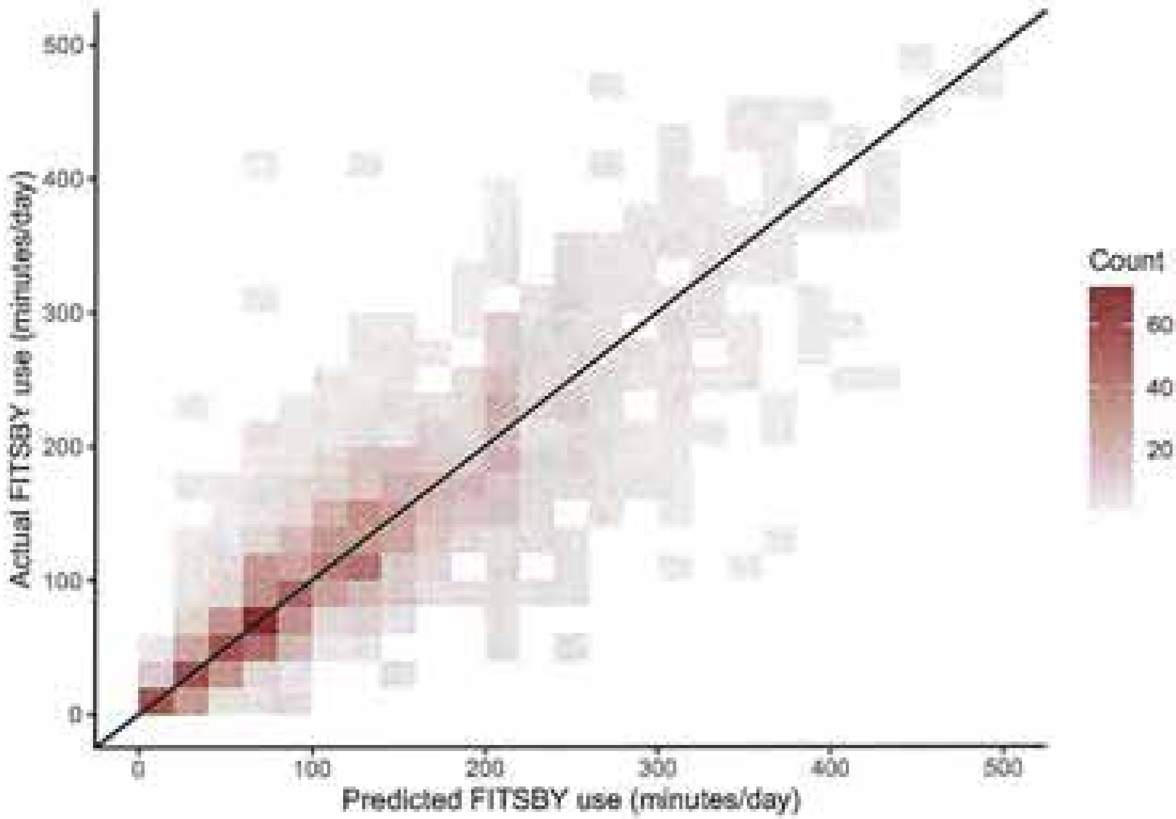
Notes: This figure presents the effects of bonus and limit treatments on self-reported change in FITSBY use on other devices relative to the three weeks before the study using equation (4). FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers, and YouTube. Self-reported changes are winsorized at 150 minutes.

D.1 Validation of Predicted Use and Multiple Price List Responses

Predicted use lines up well with actual use; see Appendix Figures A15 and A16. The \$5 (instead of \$1) prediction accuracy reward slightly reduces the absolute value of the prediction error but has tightly estimated zero effects on predicted use, actual use, and the level of the prediction error; see Appendix Table A4.

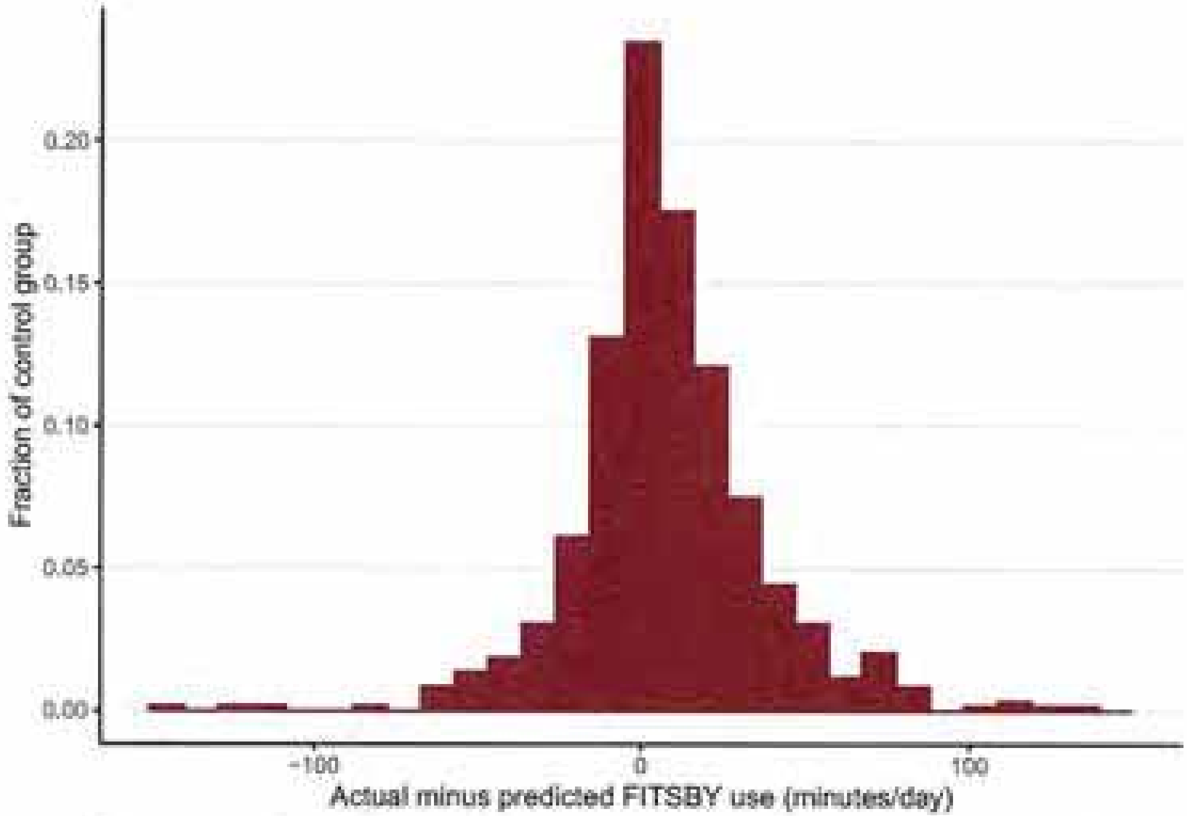
Multiple price lists are cognitively challenging, so we carry out several additional analyses to validate that these valuations are informative about people's preferences. First, participants' valuations of the bonus are correlated with the amount of money they could expect to earn; see Appendix Figure A19. Second, the limit valuation and the behavior change premium (defined in Section E.3) are correlated with each other and with *limit tighness*, *ideal use change*, *addiction scale*, *SMS addiction scale*, and other variables in expected ways; see Appendix Table A5. Third, after the bonus MPL, we asked people to "select the statement that best describes your thinking when trading off the Screen Time Bonus against the fixed payment." 24 percent responded that "I wanted to give myself an incentive to use my phone less over the next three weeks, even though it might result in a smaller payment," and this group had a substantially higher average behavior change premium; see Appendix Figures A20 and A21.

Figure A15: **Predicted vs. Actual FITSBY Use in Control**



Notes: This figure presents the number of Control group participants in each cell of actual and predicted FITSBY use across periods 2–4, using predictions from the survey just before each period. FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers, and YouTube.

Figure A16: **Histogram of Actual Minus Predicted FITSBY Use in Control Group**



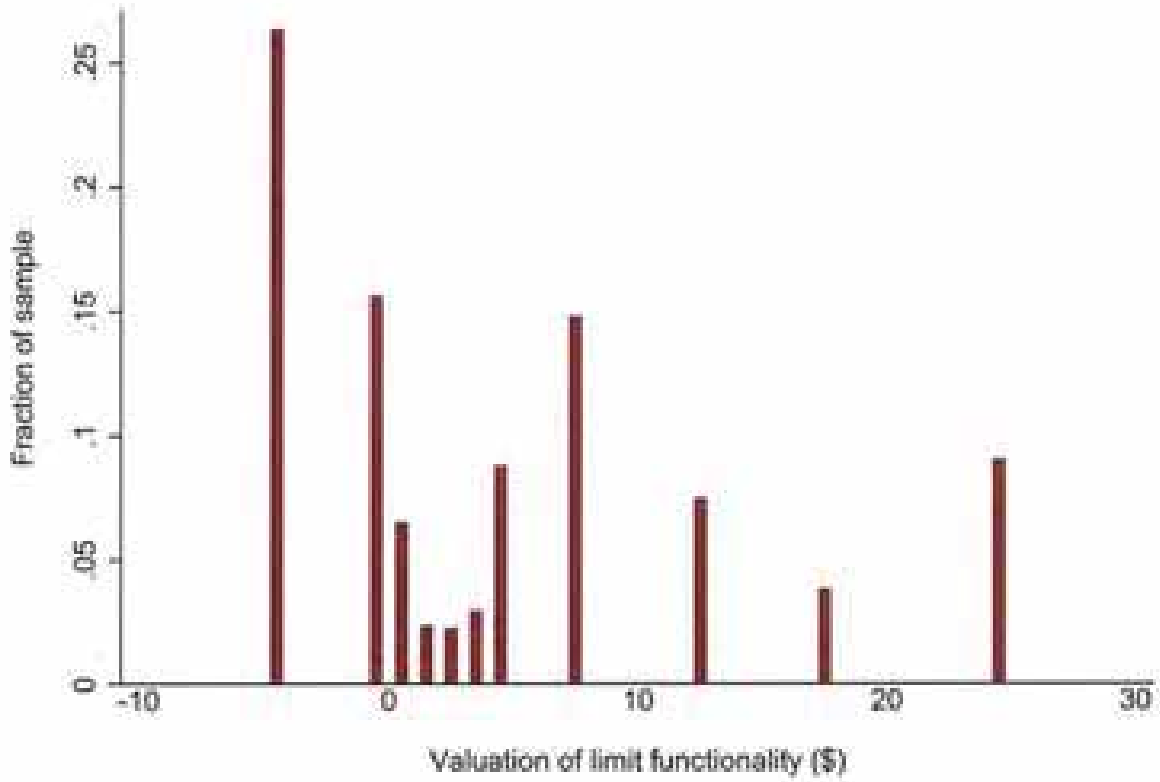
Notes: This figure presents the distribution of the difference between actual and predicted FITSBY use across periods 2–4 in the Control group, using predictions from the survey just before each period. FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers, and YouTube.

Table A4: **Effect of Prediction Accuracy Reward**

	(1)	(2)	(3)	(4)
	Predicted use	Actual use	Predicted - actual use	Absolute value of predicted - actual use
High prediction reward	1.219	3.343	-2.207	-2.379
	(2.582)	(2.386)	(1.691)	(1.435)
Constant	118.9	116.7	2.300	35.22
	(1.908)	(1.670)	(1.376)	(1.212)

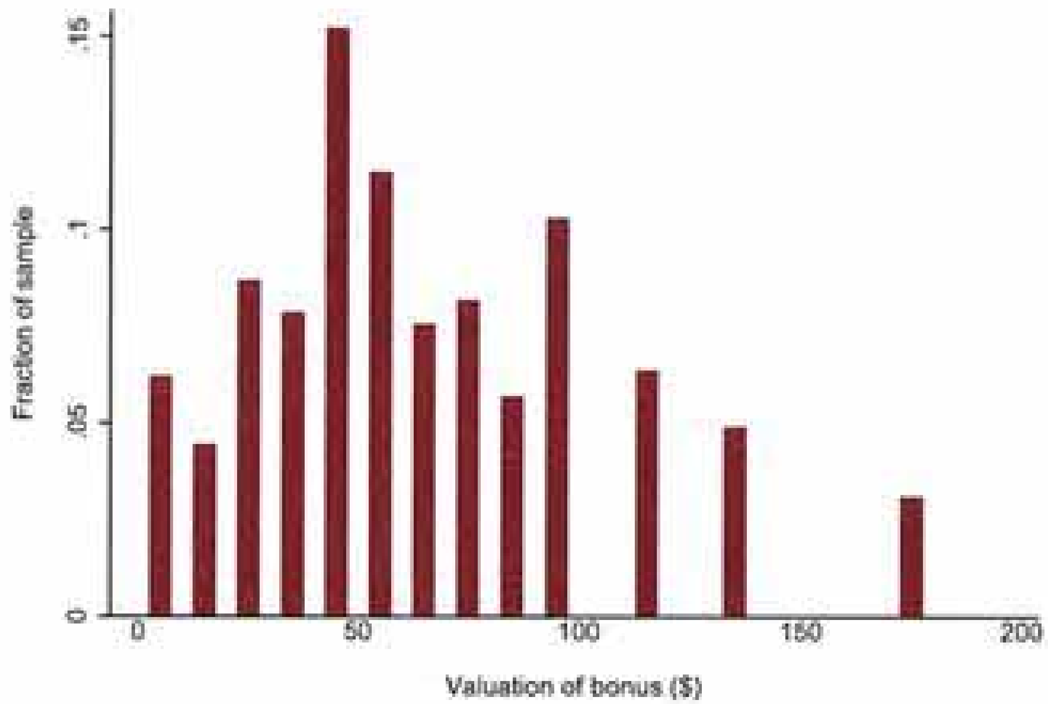
Notes: This table presents the effects of being offered the higher Prediction Reward ($5 instead of $1 for predicting within 15 minutes of actual screen time) on predicted and actual FITSBY use in minutes per day. FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers, and YouTube. Standard errors are in parentheses.

Figure A17: **Valuation of Limit Functionality**



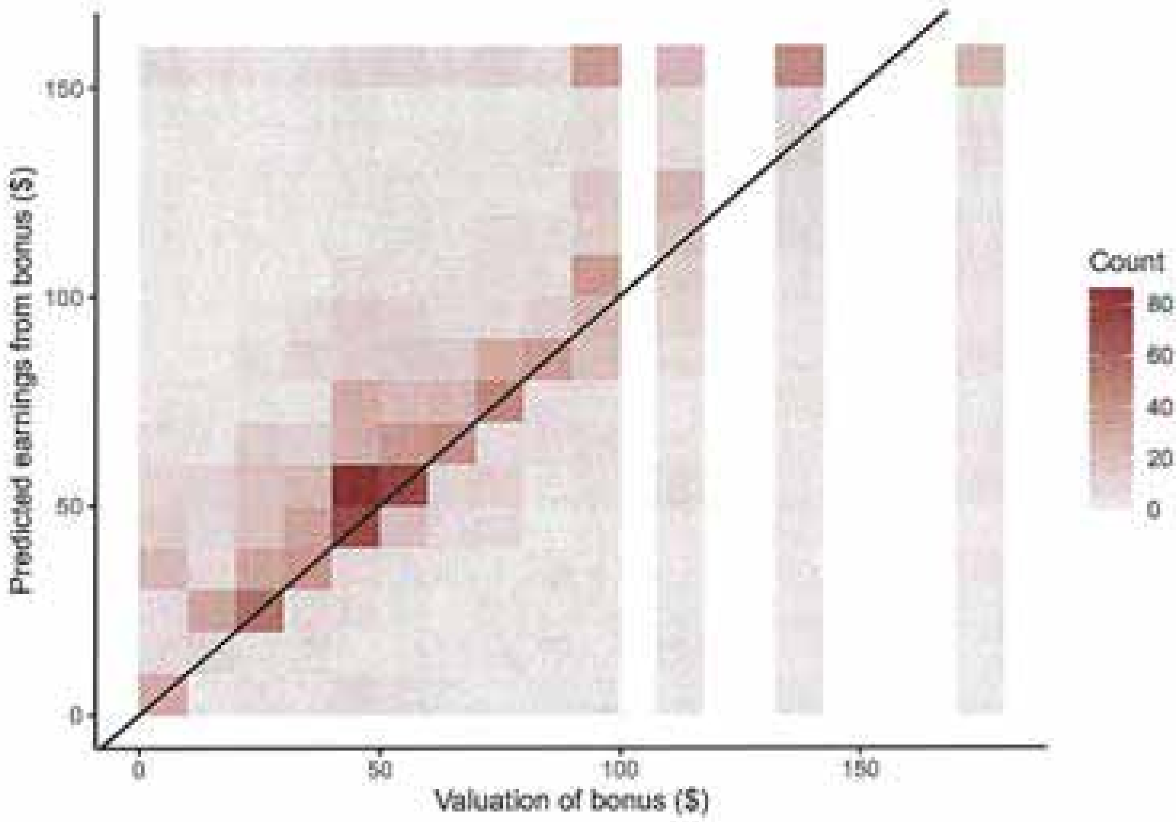
Notes: This figure presents the distribution of valuations of access to the limit functionality for the next three weeks, as elicited in a multiple price list on survey 3. Valuations above $20 are plotted at $25, and valuations below $-1 are plotted at $-5.

Figure A18: **Valuation of Screen Time Bonus**



Notes: This figure presents the distribution of valuations of the Screen Time Bonus incentive, as elicited on survey 2. Valuations above $150 are plotted at $175.

Figure A19: **Valuation of Bonus vs. Predicted Bonus Earnings**



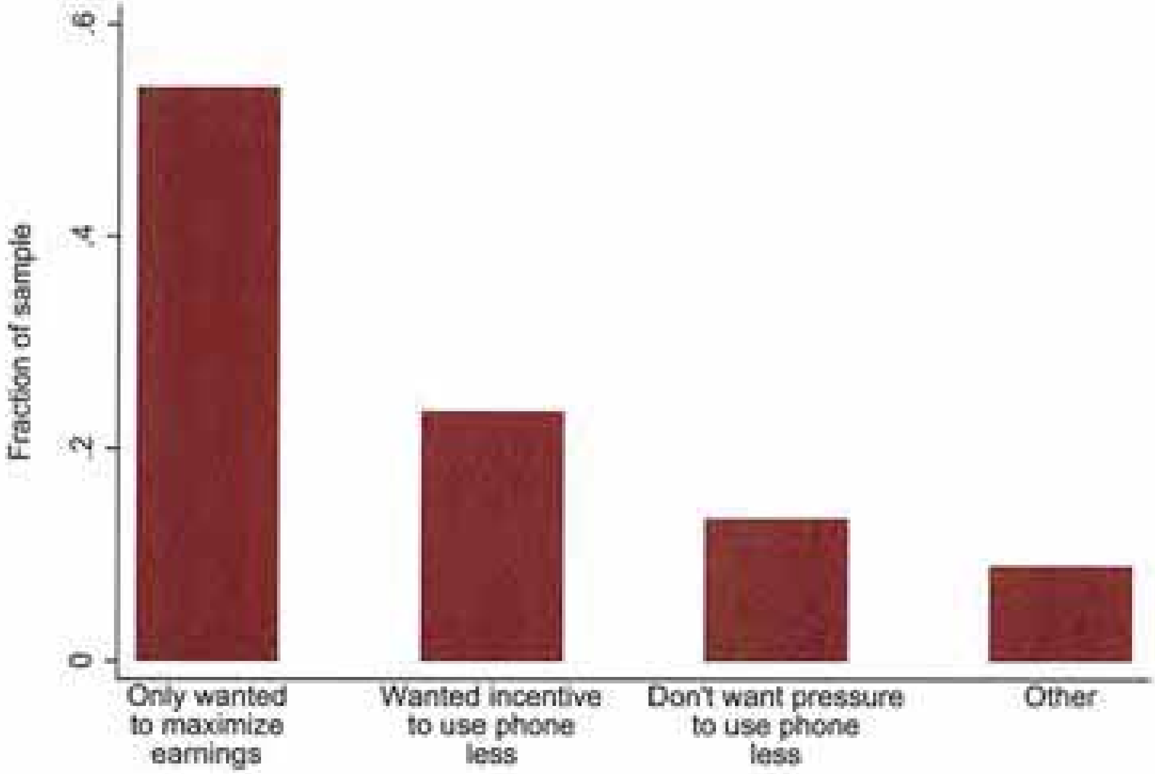
Notes: This figure presents the number of participants in each cell of predicted earnings from the Screen Time Bonus (given the participant's Bonus Benchmark and predicted FITSBY use) and valuation of the bonus, as elicited on survey 2.

Table A5: **Correlations between Temptation and Addiction Measures**

	Behavior change premium	Valuation of limit	Limit tightness	Interest in limits	Ideal use change × (-1)	Addiction scale	SMS addiction scale	Phone makes life better × (-1)
Behavior change premium	1							
Valuation of limit	0.116	1						
Limit tightness	0.471	0.199	1					
Interest in limits	0.032	0.146	0.204	1				
Ideal use change x (-1)	0.117	0.112	0.218	0.319	1			
Addiction scale	0.267	0.078	0.243	0.356	0.435	1		
SMS addiction scale	0.272	0.132	0.259	0.312	0.345	0.651	1	
Phone makes life better x (-1)	0.022	0.082	0.154	0.295	0.392	0.303	0.234	1

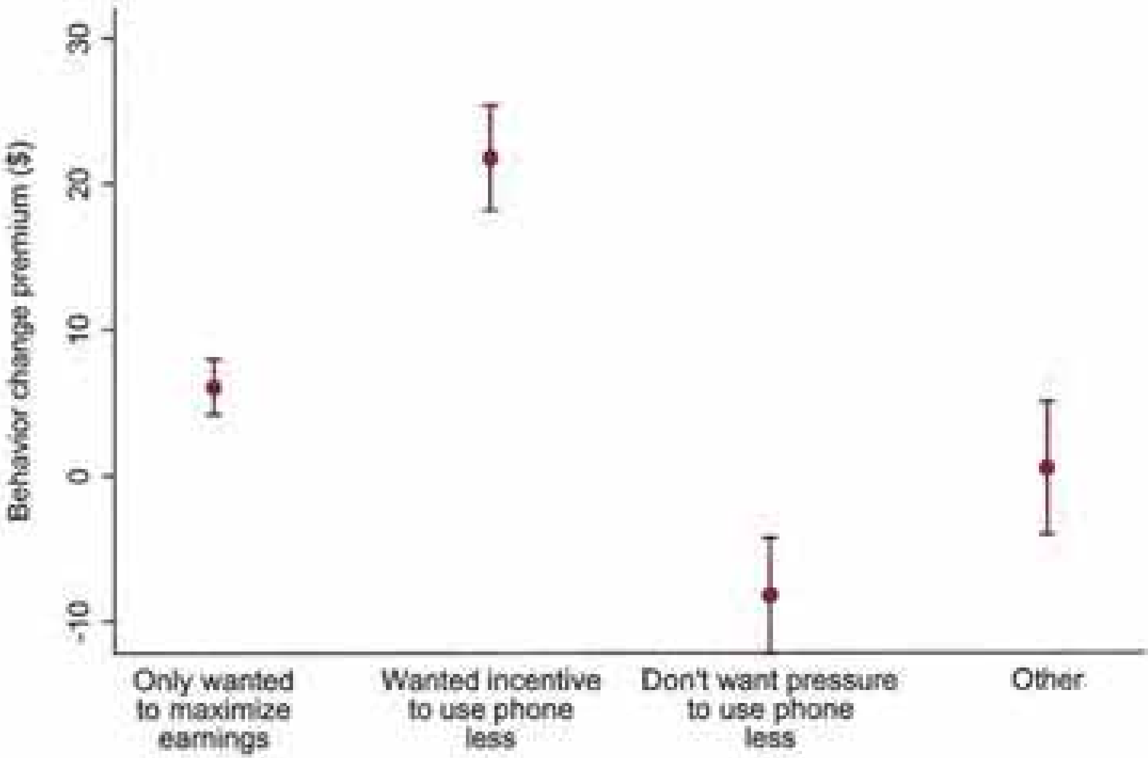
Note: The behavior change premium is the difference between the valuation of the Screen Time Bonus and the modeled valuation if the consumer believed herself to be time consistent. *Interest in limits*, *ideal use change*, *addiction scale*, *SMS addiction scale*, and *phone makes life better* are from survey 1.

Figure A20: **Reported Reasoning on Screen Time Bonus Multiple Price List**



Notes: After the bonus multiple price list, survey 2 asked participants to "select the statement that best describes your thinking when trading off the Screen Time Bonus against the fixed payment." This figure presents the share of participants who selected each answer.

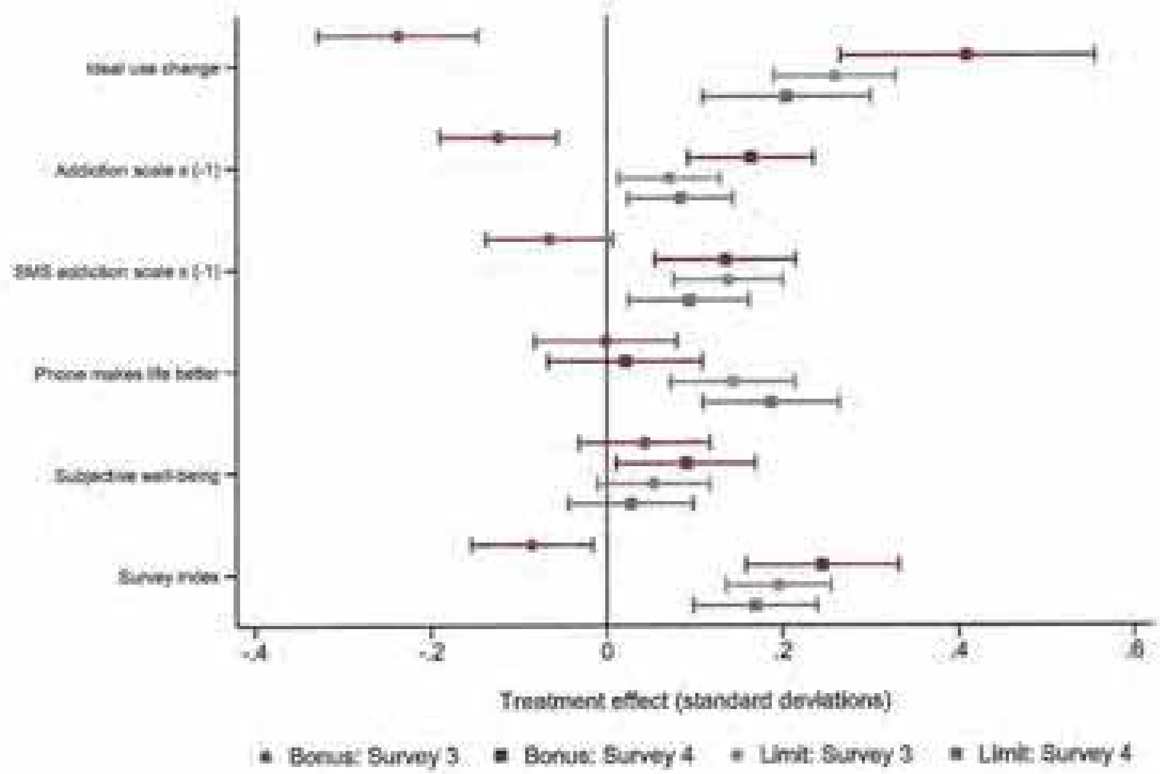
Figure A21: **Behavior Change Premium by Reported Reasoning**



Notes: The behavior change premium is the difference between the valuation of the Screen Time Bonus and the modeled valuation if the consumer believed herself to be time consistent. After the bonus multiple price list, survey 2 asked participants to "select the statement that best describes your thinking when trading off the Screen Time Bonus against the fixed payment." This figure presents means and 95 percent confidence intervals of the behavior change premium by responses to that question.

D.2 Additional Estimates of Effects on Survey Outcome Variables

Figure A22: **Effects of Limits and Bonus on Survey Outcomes on Surveys 3 and 4**



Notes: This figure presents effects of the bonus and limit treatment on survey outcome variables using equation (4), allowing separate coefficients for effects on surveys 3 vs. 4. *Ideal use change* is the answer to, "Relative to your actual use over the past 3 weeks, by how much would you ideally have [reduced/increased] your screen time?" *Addiction scale* is answers to a battery of 16 questions modified from the Mobile Phone Problem Use Scale and the Bergen Facebook Addiction Scale. *SMS addiction scale* is answers to shortened versions of the addiction scale questions delivered via text message. *Phone makes life better* is the answer to, "To what extent do you think your smartphone use made your life better or worse over the past 3 weeks?" *Subjective well-being* is answers to seven questions reflecting happiness, life satisfaction, anxiety, depression, concentration, distraction, and sleep quality; anxiety, depression, and distraction are re-oriented so that more positive reflects better subjective well-being. *Survey index* combines the previous five variables, weighting by the inverse of their covariance at baseline.

Table A6: **Treatment Effects**

(a) **Bonus**

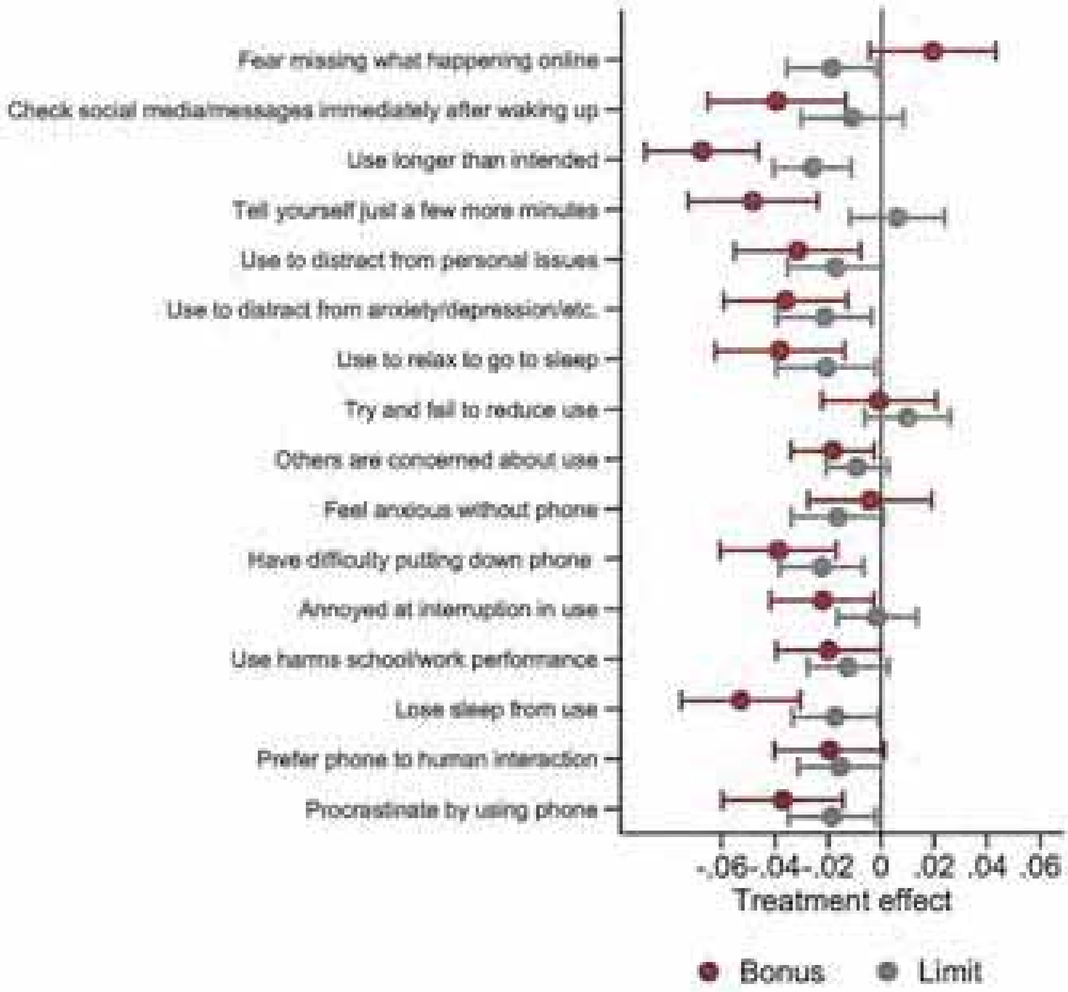
	(1) Treatment effect (original units)	(2) Standard error (original units)	(3) Treatment effect (SD units)	(4) Standard error (SD units)	(5) P-value	(6) Sharpened FDR-adjusted q-value
Ideal use change	9.0	1.6	0.41	0.074	0.000	0.000
Addiction scale x (-1)	0.44	0.10	0.16	0.037	0.000	0.000
SMS addiction scale x (-1)	0.42	0.13	0.14	0.041	0.001	0.004
Phone makes life better	0.042	0.090	0.021	0.045	0.64	0.78
Subjective well-being	0.23	0.10	0.090	0.040	0.026	0.09
Survey index	0.17	0.031	0.24	0.044	0.000	0.000

(b) **Limit**

	(1) Treatment effect (original units)	(2) Standard error (original units)	(3) Treatment effect (SD units)	(4) Standard error (SD units)	(5) P-value	(6) Sharpened FDR-adjusted q-value
Ideal use change	5.1	0.75	0.23	0.034	0.000	0.000
Addiction scale x (-1)	0.21	0.071	0.078	0.027	0.004	0.008
SMS addiction scale x (-1)	0.36	0.090	0.12	0.028	0.000	0.000
Phone makes life better	0.33	0.064	0.16	0.032	0.000	0.000
Subjective well-being	0.10	0.075	0.040	0.030	0.18	0.24
Survey index	0.13	0.020	0.18	0.029	0.000	0.000

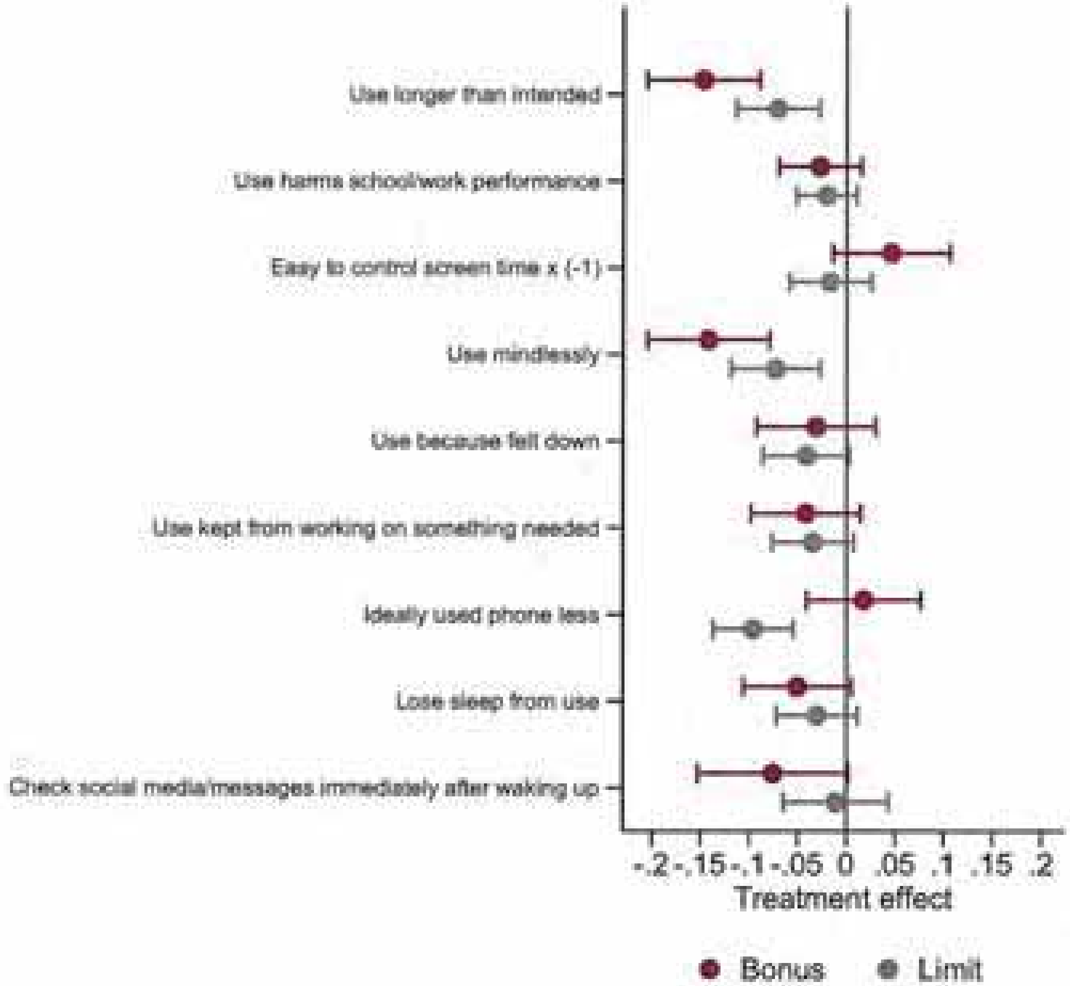
Notes: This table presents effects of the bonus and limit treatments on survey outcome variables using equation (4). The bonus effect is measured on survey 4, while the limit effect is measured on both surveys 3 and 4. *Ideal use change* is the answer to, "Relative to your actual use over the past 3 weeks, by how much would you ideally have [reduced/increased] your screen time?" *Addiction scale* is answers to a battery of 16 questions modified from the Mobile Phone Problem Use Scale and the Bergen Facebook Addiction Scale. *SMS addiction scale* is answers to shortened versions of the addiction scale questions delivered via text message. *Phone makes life better* is the answer to, "To what extent do you think your smartphone use made your life better or worse over the past 3 weeks?" *Subjective well-being* is answers to seven questions reflecting happiness, life satisfaction, anxiety, depression, concentration, distraction, and sleep quality; anxiety, depression, and distraction are re-oriented so that more positive reflects better subjective well-being. *Survey index* combines the previous five variables, weighting by the inverse of their covariance at baseline. The effects in standard deviation units in column 3 match those reported on Figure 8.

Figure A23: **Effects on Addiction Responses**



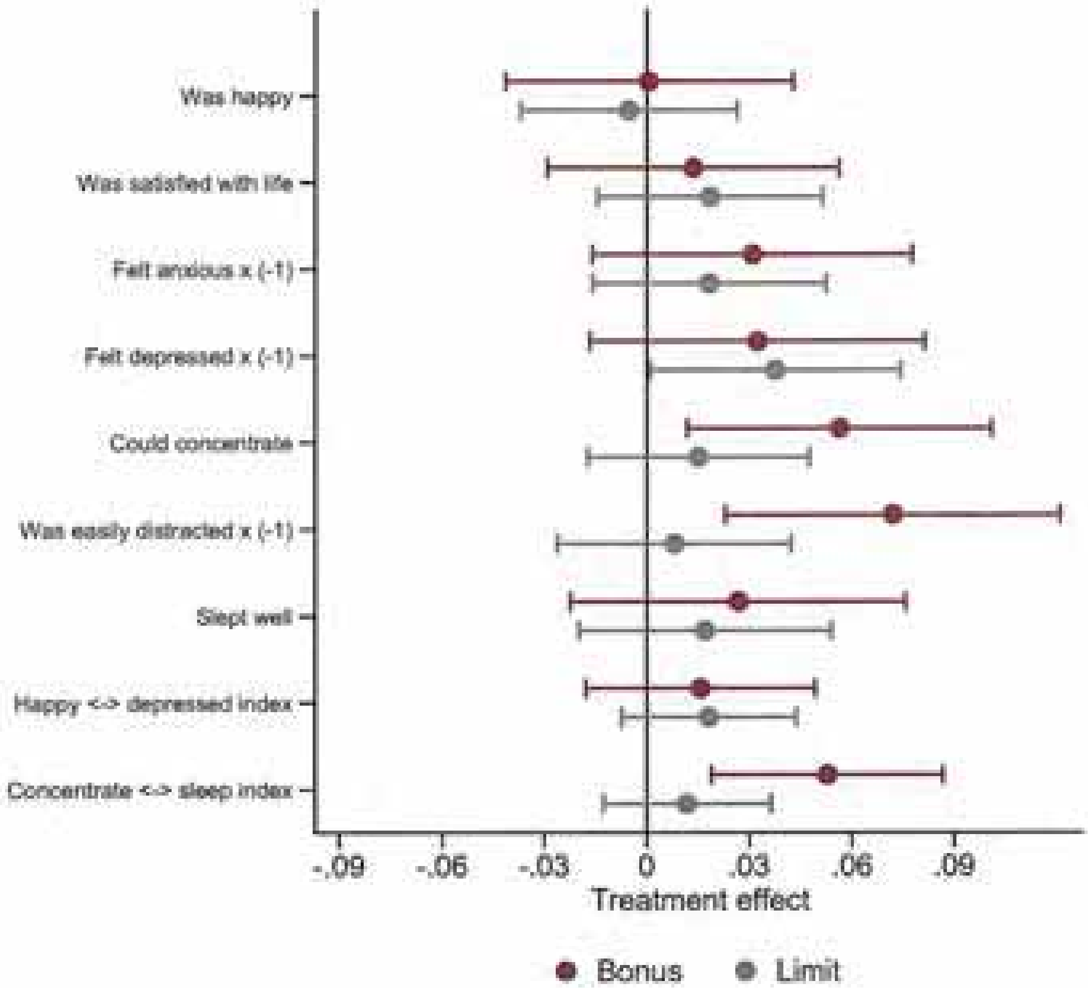
Notes: This figure presents the effects of the bonus and limit treatments on individual items in the *addiction scale* variable using equation (4). The bonus effect is measured on survey 4, while the limit effect is measured on both surveys 3 and 4. The direction of the effects in this figure are opposite those in the main figures, because *addiction scale* is multiplied by -1 in those figures.

Figure A24: **Effects on SMS Addiction Responses**



Notes: This figure presents the effects of the bonus and limit treatments on individual items in the *SMS addiction scale* variable using equation (4). The bonus effect is measured on survey 4, while the limit effect is measured on both surveys 3 and 4. The direction of the effects in this figure are opposite those in the main figures, because *SMS addiction scale* is multiplied by -1 in those figures.

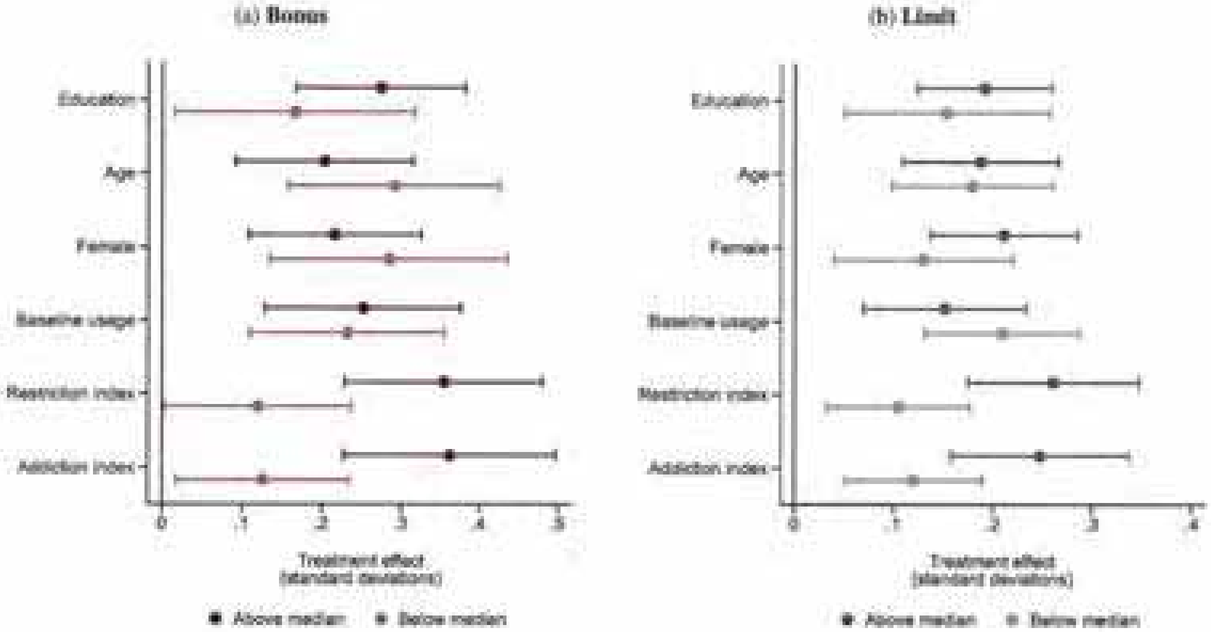
Figure A25: **Effects on Subjective Well-Being Responses**



Notes: This figure presents the effects of the bonus and limit treatments on individual items in the *subjective well-being* variable using equation (4). The bonus effect is measured on survey 4, while the limit effect is measured on both surveys 3 and 4.

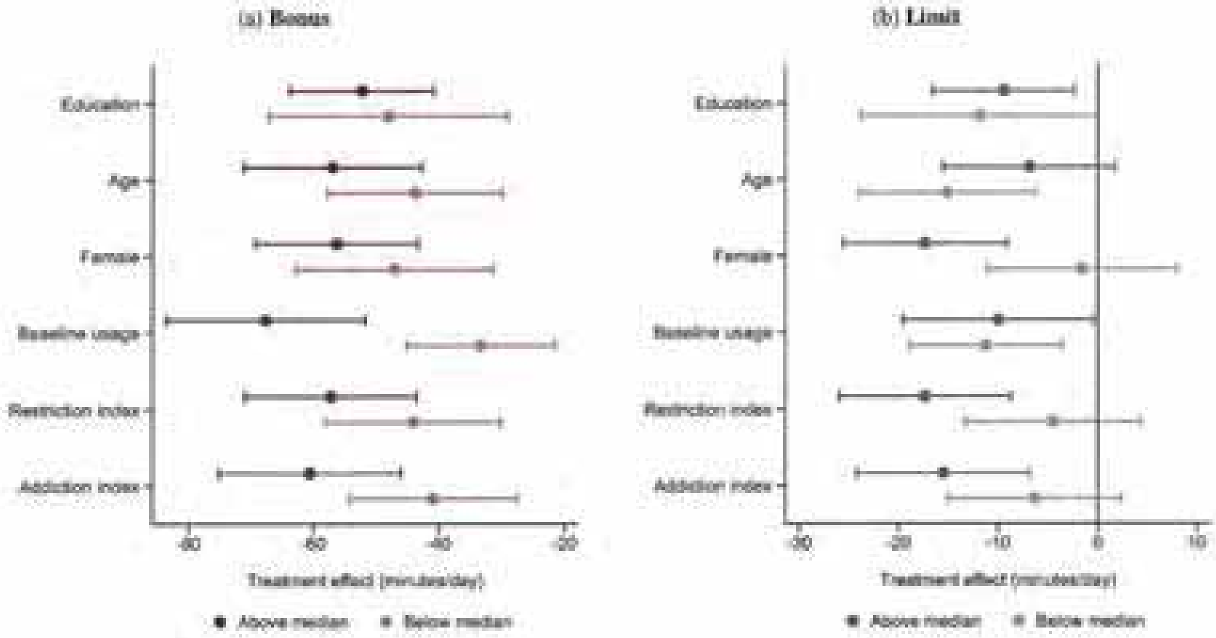
D.3 Heterogeneous Treatment Effects

Figure A26: **Heterogeneous Effects of Limits and Bonus on Survey Index**



Notes: This figure presents heterogeneous effects of the bonus and limit treatments on *survey index*, the inverse-covariance weighted average of five measures of smartphone addiction and subjective well-being, using equation (4). The bonus effect is measured on survey 4, while the limit effect is measured on both surveys 3 and 4. Above-median education includes people with a college degree or more, above-median age includes people 30 and older, and median baseline FITSBY use is 137 minutes per day. *Restriction index* is a combination of *interest in limits* and *ideal use change*. *Addiction index* is a combination of *addiction scale* and *phone makes life better*.

Figure A27: **Heterogeneous Effects of Limits and Bonus on FITSBY Use**



Notes: This figure presents heterogeneous effects of the bonus and limit treatments on FITSBY use using equation (4). The bonus effects are measured in period 3, while the limit effects are measured in periods 2–5. FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers, and YouTube. Above-median education includes people with a college degree or more, above-median age includes people 30 and older, and median baseline FITSBY use is 137 minutes per day. *Restriction index* is a combination of *interest in limits* and *ideal use change*. *Addiction index* is a combination of *addiction scale* and *phone makes life better*.

D.4 Local Average Treatment Effects on Survey Outcomes

Our pre-analysis plan specified that we would also estimate instrumental variables (IV) regressions with previous period FITSBY use $x_{i,t-1}$ as the endogenous variable:

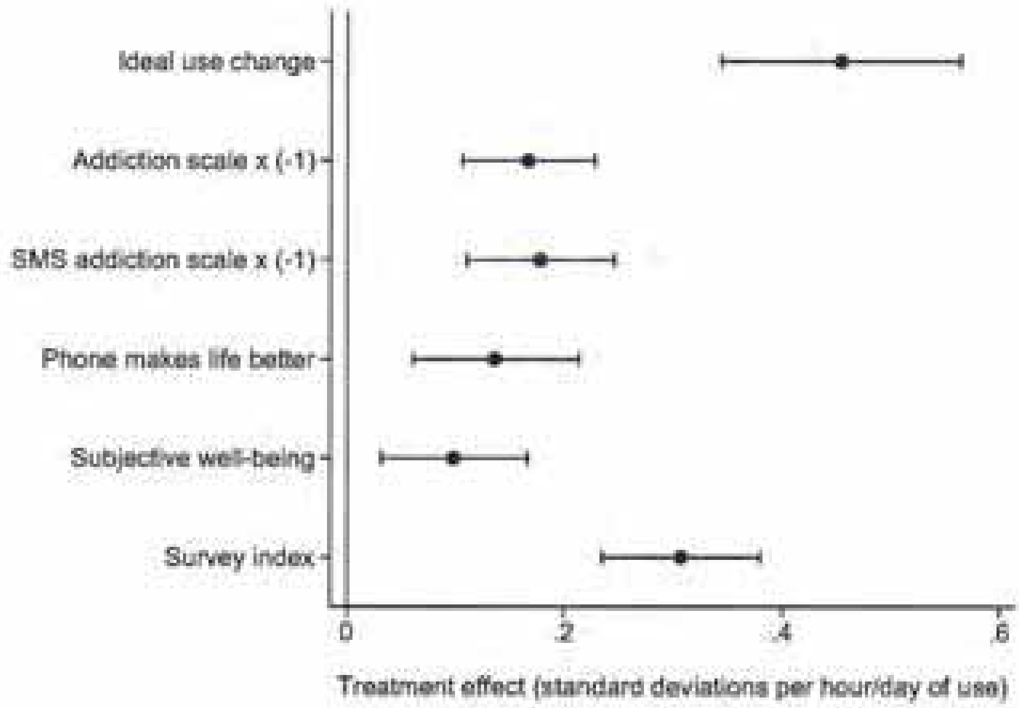
$$Y_{it} = \tau x_{i,t-1} + \beta_t X_{i1} + \nu_{it} + \varepsilon_i, \quad (17)$$

instrumenting for $x_{i,t-1}$ with B_i and L_i interacted with $t = 3$ and $t = 4$ indicators. We combine data from surveys 3 and 4 and let all coefficients other than τ vary across the two periods. Conceptually, this regression combines the effects of the bonus and limit intervention, weighting the interventions by their effects on FITSBY use. Because the limit treatment could affect survey outcomes through channels other than reduced FITSBY use—for example, by giving people an increased feeling of control over their screen time—we do not claim that the IV exclusion restriction necessarily holds.

Appendix Figure A28 presents local average treatment effects estimated using equation (17), combining effects from both treatments. Appendix Figures A29–A34 study heterogeneity along the six pre-specified

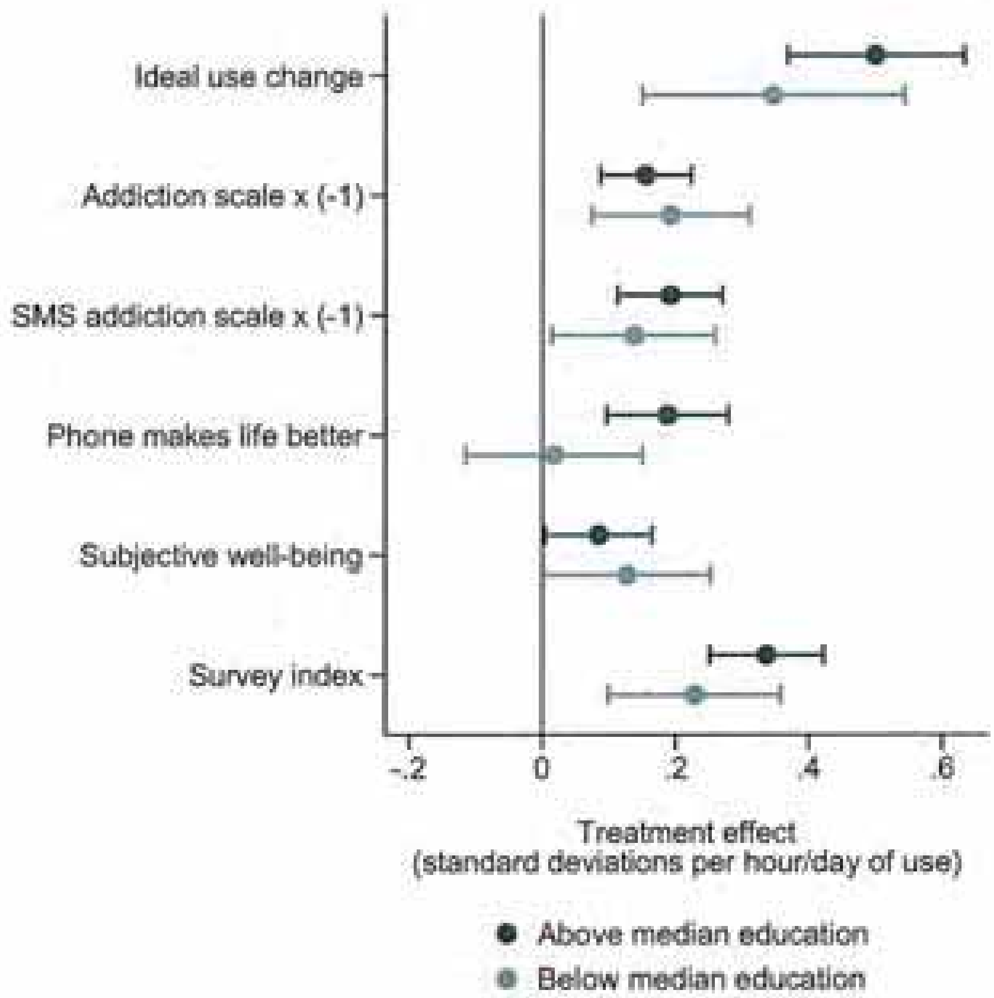
moderators. The results are qualitatively similar to Figures 8 and A26, except that the estimates are slightly more precise, as would be expected from combining effects of two interventions. Note that since the average effects of both interventions are about the same for people with low versus high baseline use (Figure A26), the local average treatment effects of reduced use are much larger for people with low baseline use (Appendix Figure A32).

Figure A28: **Local Average Treatment Effects of FITSBY Use on Survey Outcome Variables**



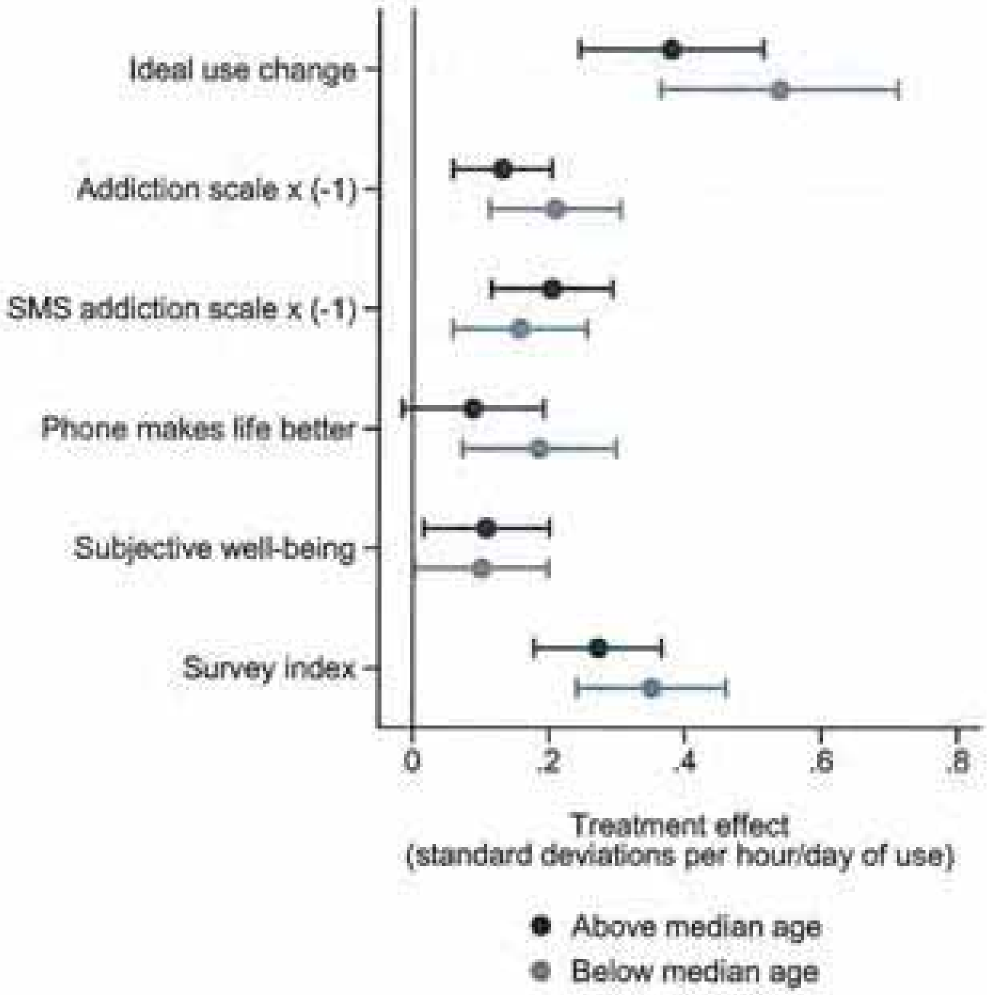
Notes: This figure presents local average treatment effects of FITSBY use on survey outcome variables using equation (17). We instrument for FITSBY use with Bonus and Limit group indicators interacted with period indicators. FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers, and YouTube. *Ideal use change* is the answer to, "Relative to your actual use over the past 3 weeks, by how much would you ideally have [reduced/increased] your screen time?" *Addiction scale* is answers to a battery of 16 questions modified from the Mobile Phone Problem Use Scale and the Bergen Facebook Addiction Scale. *SMS addiction scale* is answers to shortened versions of the addiction scale questions delivered via text message. *Phone makes life better* is the answer to, "To what extent do you think your smartphone use made your life better or worse over the past 3 weeks?" *Subjective well-being* is answers to seven questions reflecting happiness, life satisfaction, anxiety, depression, concentration, distraction, and sleep quality; anxiety, depression, and distraction are re-oriented so that more positive reflects better subjective well-being. *Survey index* combines the previous five variables, weighting by the inverse of their covariance at baseline.

Figure A29: **Heterogeneous Effects on Survey Outcome Variables by Education**



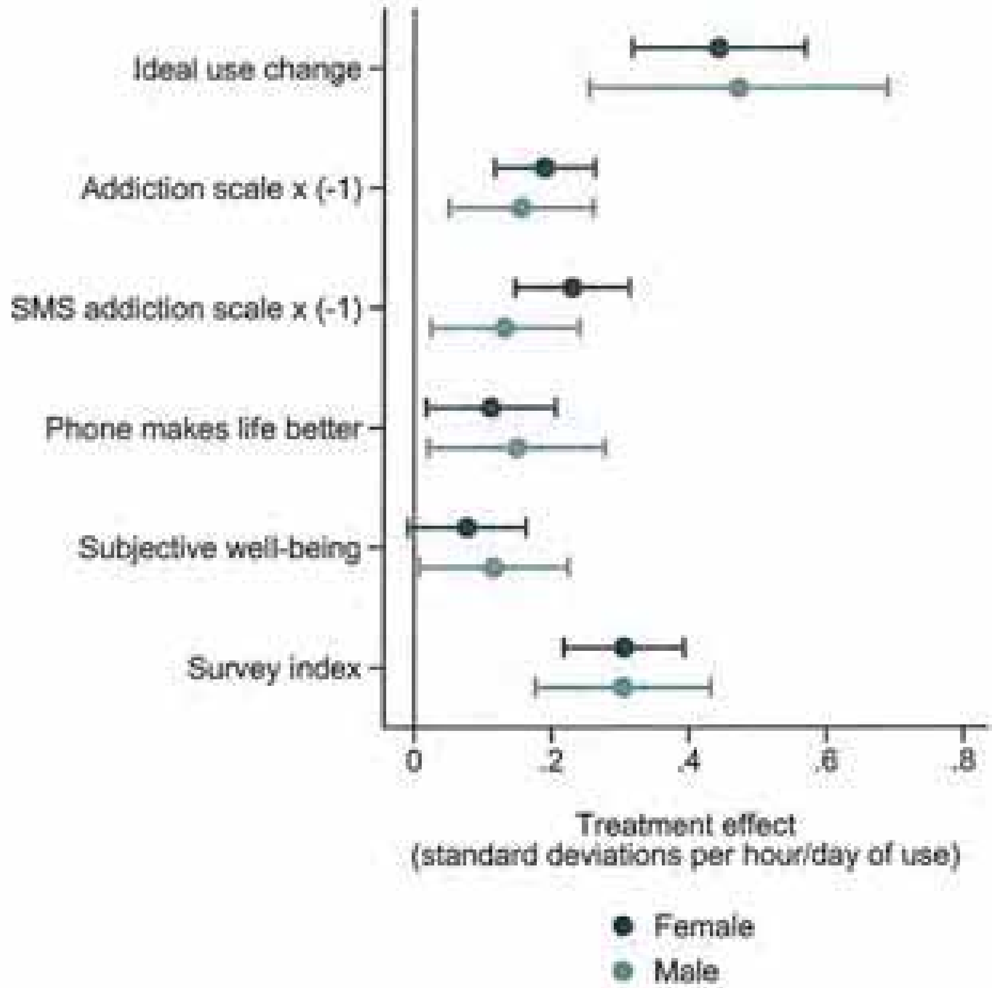
Notes: This figure presents local average treatment effects of FITSBY use on survey outcome variables using equation (17), for above- and below-median education. We instrument for FITSBY use with Bonus and Limit group indicators interacted with period indicators. FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers, and YouTube. *Ideal use change* is the answer to, "Relative to your actual use over the past 3 weeks, by how much would you ideally have [reduced/increased] your screen time?" *Addiction scale* is answers to a battery of 16 questions modified from the Mobile Phone Problem Use Scale and the Bergen Facebook Addiction Scale. *SMS addiction scale* is answers to shortened versions of the addiction scale questions delivered via text message. *Phone makes life better* is the answer to, "To what extent do you think your smartphone use made your life better or worse over the past 3 weeks?" *Subjective well-being* is answers to seven questions reflecting happiness, life satisfaction, anxiety, depression, concentration, distraction, and sleep quality; anxiety, depression, and distraction are re-oriented so that more positive reflects better subjective well-being. *Survey index* combines the previous five variables, weighting by the inverse of their covariance at baseline.

Figure A30: **Heterogeneous Effects on Survey Outcome Variables by Age**



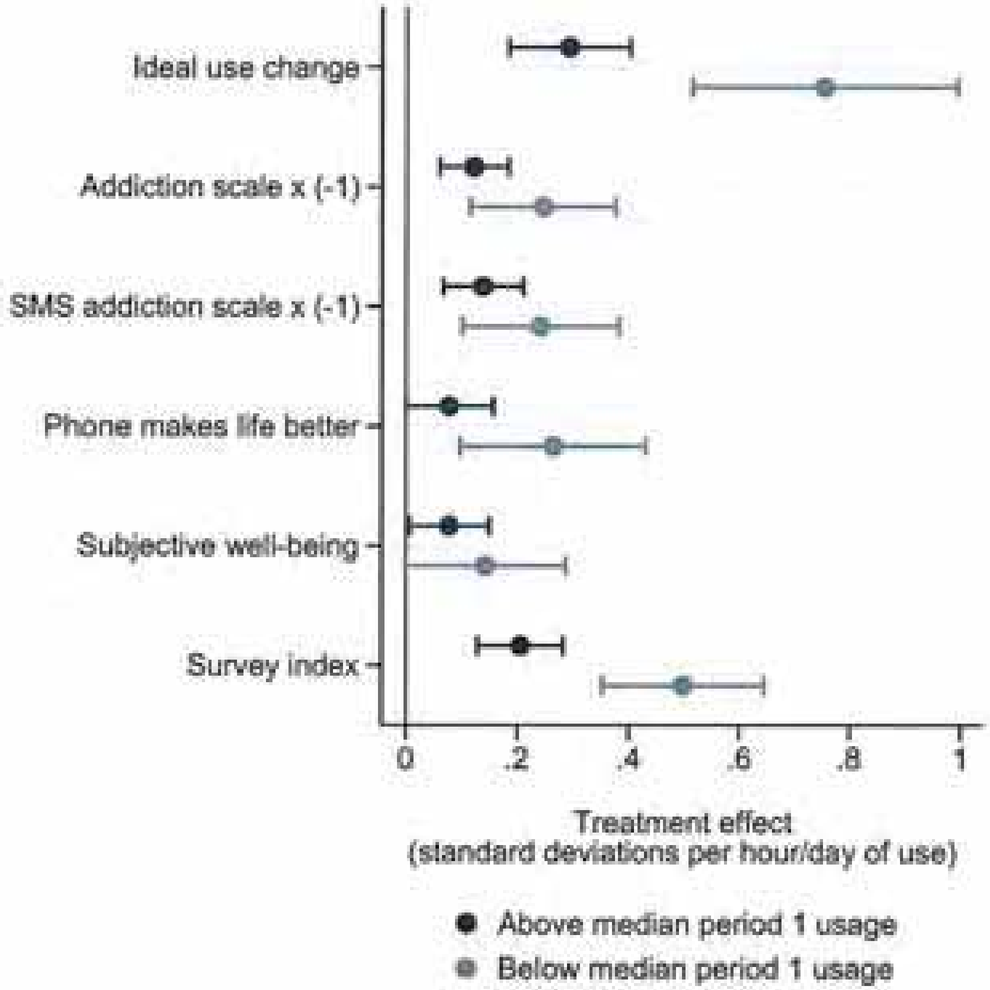
Notes: This figure presents local average treatment effects of FITSBY use on survey outcome variables using equation (17), for above- and below-median age. We instrument for FITSBY use with Bonus and Limit group indicators interacted with period indicators. FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers, and YouTube. *Ideal use change* is the answer to, "Relative to your actual use over the past 3 weeks, by how much would you ideally have [reduced/increased] your screen time?" *Addiction scale* is answers to a battery of 16 questions modified from the Mobile Phone Problem Use Scale and the Bergen Facebook Addiction Scale. *SMS addiction scale* is answers to shortened versions of the addiction scale questions delivered via text message. *Phone makes life better* is the answer to, "To what extent do you think your smartphone use made your life better or worse over the past 3 weeks?" *Subjective well-being* is answers to seven questions reflecting happiness, life satisfaction, anxiety, depression, concentration, distraction, and sleep quality; anxiety, depression, and distraction are re-oriented so that more positive reflects better subjective well-being. *Survey index* combines the previous five variables, weighting by the inverse of their covariance at baseline.

Figure A31: **Heterogeneous Effects on Survey Outcome Variables by Gender**



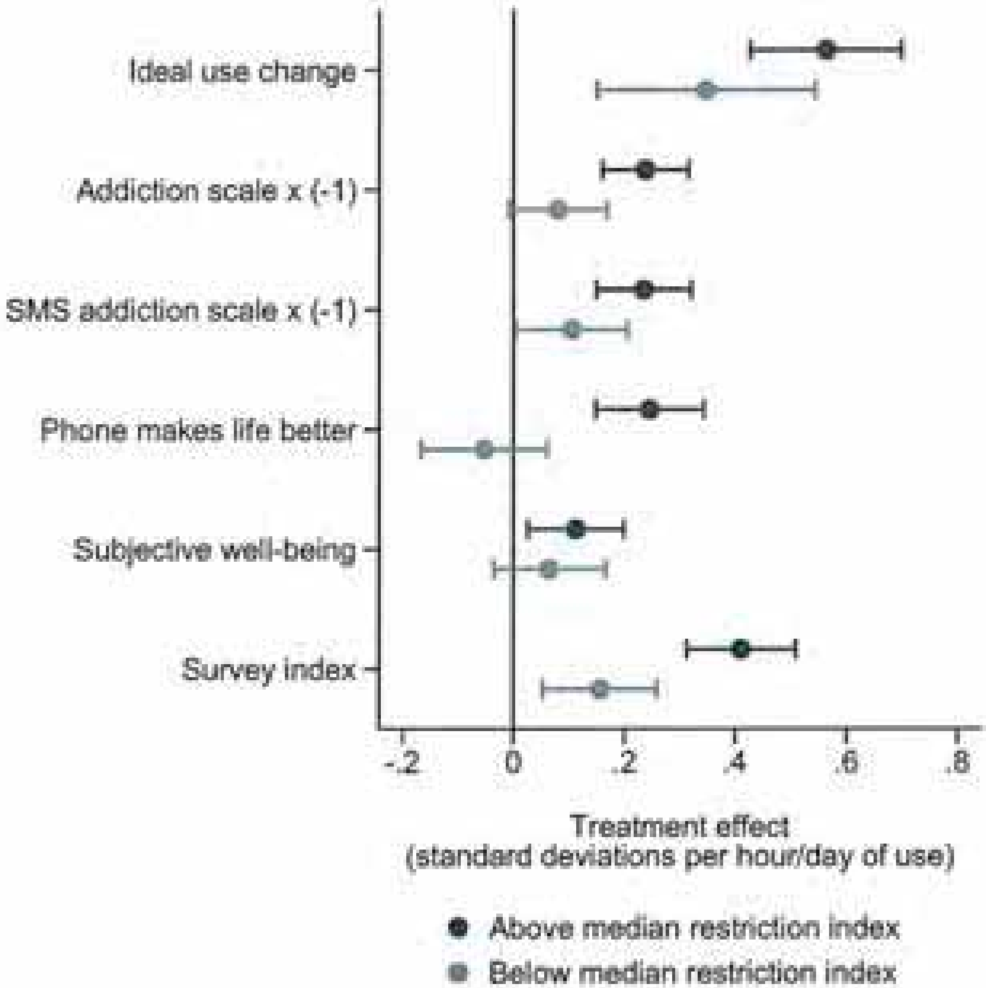
Notes: This figure presents local average treatment effects of FITSBY use on survey outcome variables using equation (17), for men versus women. We instrument for FITSBY use with Bonus and Limit group indicators interacted with period indicators. FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers, and YouTube. *Ideal use change* is the answer to, "Relative to your actual use over the past 3 weeks, by how much would you ideally have [reduced/increased] your screen time?" *Addiction scale* is answers to a battery of 16 questions modified from the Mobile Phone Problem Use Scale and the Bergen Facebook Addiction Scale. *SMS addiction scale* is answers to shortened versions of the addiction scale questions delivered via text message. *Phone makes life better* is the answer to, "To what extent do you think your smartphone use made your life better or worse over the past 3 weeks?" *Subjective well-being* is answers to seven questions reflecting happiness, life satisfaction, anxiety, depression, concentration, distraction, and sleep quality; anxiety, depression, and distraction are re-oriented so that more positive reflects better subjective well-being. *Survey index* combines the previous five variables, weighting by the inverse of their covariance at baseline.

Figure A32: **Heterogeneous Effects on Survey Outcome Variables by Baseline FITSBY Use**



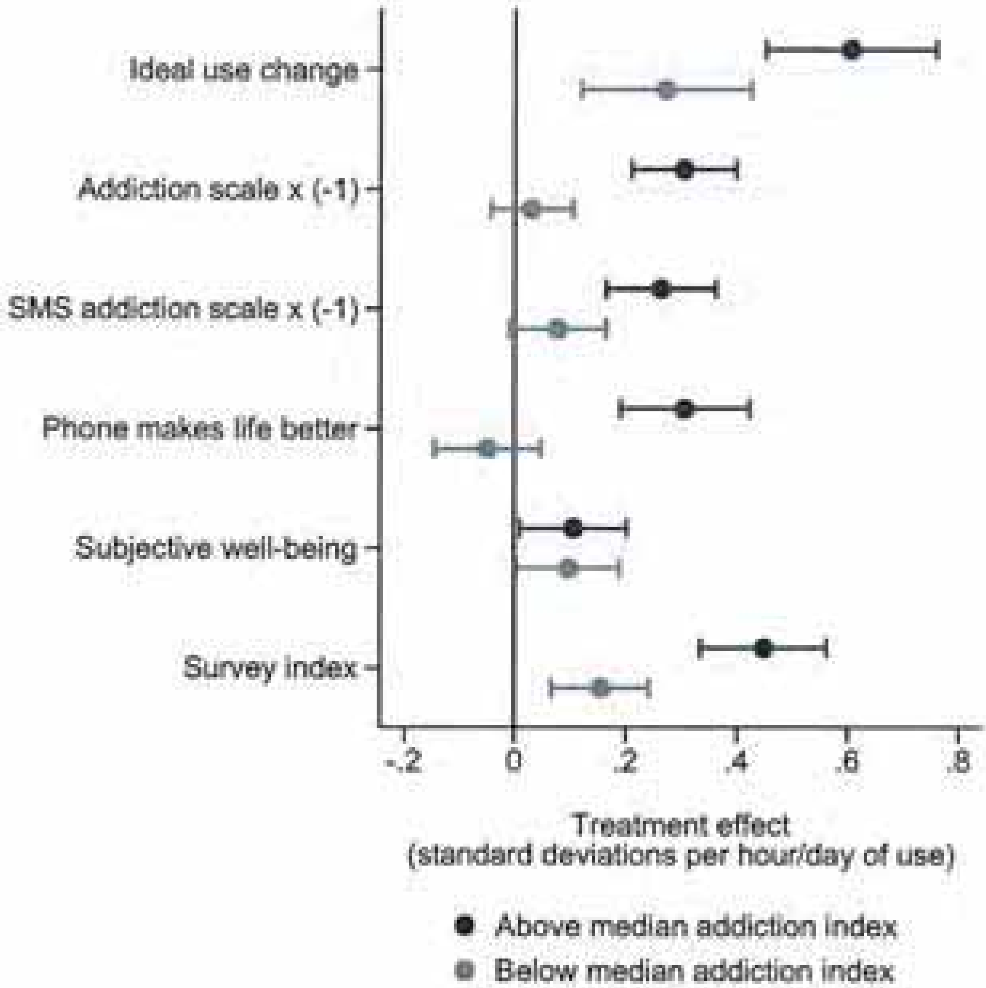
Notes: This figure presents local average treatment effects of FITSBY use on survey outcome variables using equation (17), for above- and below-median baseline FITSBY use. We instrument for FITSBY use with Bonus and Limit group indicators interacted with period indicators. FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers, and YouTube. *Ideal use change* is the answer to, "Relative to your actual use over the past 3 weeks, by how much would you ideally have [reduced/increased] your screen time?" *Addiction scale* is answers to a battery of 16 questions modified from the Mobile Phone Problem Use Scale and the Bergen Facebook Addiction Scale. *SMS addiction scale* is answers to shortened versions of the addiction scale questions delivered via text message. *Phone makes life better* is the answer to, "To what extent do you think your smartphone use made your life better or worse over the past 3 weeks?" *Subjective well-being* is answers to seven questions reflecting happiness, life satisfaction, anxiety, depression, concentration, distraction, and sleep quality; anxiety, depression, and distraction are re-oriented so that more positive reflects better subjective well-being. *Survey index* combines the previous five variables, weighting by the inverse of their covariance at baseline.

Figure A33: **Heterogeneous Effects on Survey Outcome Variables by Restriction Index**



Notes: This figure presents local average treatment effects of FITSBY use on survey outcome variables using equation (17), for above- and below-median values of *restriction index*, a combination of *interest in limits* and *ideal use change*. We instrument for FITSBY use with Bonus and Limit group indicators interacted with period indicators. FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers, and YouTube. *Ideal use change* is the answer to, "Relative to your actual use over the past 3 weeks, by how much would you ideally have [reduced/increased] your screen time?" *Addiction scale* is answers to a battery of 16 questions modified from the Mobile Phone Problem Use Scale and the Bergen Facebook Addiction Scale. *SMS addiction scale* is answers to shortened versions of the addiction scale questions delivered via text message. *Phone makes life better* is the answer to, "To what extent do you think your smartphone use made your life better or worse over the past 3 weeks?" *Subjective well-being* is answers to seven questions reflecting happiness, life satisfaction, anxiety, depression, concentration, distraction, and sleep quality; anxiety, depression, and distraction are re-oriented so that more positive reflects better subjective well-being. *Survey index* combines the previous five variables, weighting by the inverse of their covariance at baseline.

Figure A34: **Heterogeneous Effects on Survey Outcome Variables by Addiction Index**



Notes: This figure presents local average treatment effects of FITSBY use on survey outcome variables using equation (17), for above- and below-median values of *addiction index*, a combination of *addiction scale* and *phone makes life better*. We instrument for FITSBY use with Bonus and Limit group indicators interacted with period indicators. FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browsers, and YouTube. *Ideal use change* is the answer to, "Relative to your actual use over the past 3 weeks, by how much would you ideally have [reduced/increased] your screen time?" *Addiction scale* is answers to a battery of 16 questions modified from the Mobile Phone Problem Use Scale and the Bergen Facebook Addiction Scale. *SMS addiction scale* is answers to shortened versions of the addiction scale questions delivered via text message. *Phone makes life better* is the answer to, "To what extent do you think your smartphone use made your life better or worse over the past 3 weeks?" *Subjective well-being* is answers to seven questions reflecting happiness, life satisfaction, anxiety, depression, concentration, distraction, and sleep quality; anxiety, depression, and distraction are re-oriented so that more positive reflects better subjective well-being. *Survey index* combines the previous five variables, weighting by the inverse of their covariance at baseline.

E Unrestricted Model and Alternative Temptation Estimates

In this appendix, we estimate the unrestricted model and present alternative estimates of the temptation parameter γ.

E.1 Key Theoretical Results

Three theoretical results are key to our estimation strategy: the Euler equation, linear policy functions, and the steady state.

Euler equation. The first-order conditions of equation (2) for periods t and $t+1$ can be re-arranged into an Euler equation characterizing the equilibrium relationship between consumption in periods t and $t+1$. To simplify notation, define $u_t := u_t(x_t^*; s_t, p_t)$ as current utility, define $\tilde{x}_r := \tilde{x}_r^*(\tilde{s}_r, \tilde{\gamma}, \mathbf{p}_r)$ and $\tilde{u}_r := u_r(\tilde{x}_r; \tilde{s}_r, p_r)$ as predicted consumption and utility for future periods $r > t$, and define $\tilde{\lambda}_r := \frac{\partial \tilde{x}_r}{\partial \tilde{s}_r}$ as the predicted effect of habit stock on consumption.

Proposition 1. *Suppose $u_t(x_t; s_t, p_t)$ is given by equation (3) and $(x_0^*, ..., x_T^*)$ is a perception-perfect strategy profile with differentiable strategies. Then for each $t < T$,*

$$\underbrace{\eta x_t^* + \zeta s_t + \xi_t - p_t}_{\partial u_t / \partial x_t} + \gamma = (1-\alpha)\delta\rho \left[\underbrace{\eta \tilde{x}_{t+1} + \zeta \tilde{s}_{t+1} + \xi_{t+1} - p_{t+1}}_{\partial \tilde{u}_{t+1} / \partial \tilde{x}_{t+1}} + \tilde{\gamma} + \tilde{\gamma}\tilde{\lambda}_{t+1} - \underbrace{(\zeta \tilde{x}_{t+1} + \phi)}_{\partial \tilde{u}_{t+1} / \partial \tilde{s}_{t+1}} \right]. \quad (18)$$

Proof. See Appendix F.1.

□

With full myopia ($\delta = 0$) or full projection bias ($\alpha = 1$), consumers maximize current-period flow utility, setting the left-hand side of equation (18) to zero. In a "rational" habit formation model with $\alpha = 0$ and $\tilde{\gamma} = \gamma = 0$, the right-hand side adds two effects. First, there is an adjacent complementarity effect where people consume more in period t (driving down marginal utility $\partial u_t / \partial x_t$) if they expect to consume more in $t+1$ (i.e. if future marginal utility $\partial \tilde{u}_{t+1} / \partial x_{t+1}$ is lower). Second, there is a direct habit stock effect where people consume more in period t if the marginal utility from the resulting habit stock $\partial \tilde{u}_{t+1} / \partial s_{t+1}$ is higher.

Temptation adds two forces. First, the balance of the adjacent complementarity effect tilts toward increased consumption, as γ is added to period t marginal utility and $\tilde{\gamma}$ is added to predicted period $t+1$ marginal utility. Second, people reduce current consumption to avoid exacerbating perceived future overconsumption, giving $\tilde{\gamma}\lambda_{t+1}$ on the right-hand side.

Linear policy functions. With quadratic flow utility, equilibrium consumption is linear in habit stock with slope λ_t, and equilibrium predicted consumption is linear in habit stock with slope $\tilde{\lambda}_t$. Furthermore, if

the consumer's objective function is concave, λ and $\bar{\lambda}$ are constant far from the time horizon. This argument follows Gruber and Köszegi (2001).

Proposition 2. *Suppose the conditions for Proposition 1 hold. Then for any t,*

$$x_t^*(s_t, \gamma, p_t) = \lambda_t s_t + \mu_t(\gamma) \tag{19}$$
$$\tilde{x}_t^*(s_t, \tilde{\gamma}, p_t) = \bar{\lambda}_t s_t + \mu_t(\tilde{\gamma}), \tag{20}$$

where λ_t *is a function of only* $\{\eta, \zeta, \delta, \rho, \alpha\}$*,* $\bar{\lambda}_t$ *is a function of only* $\{\eta, \zeta, \delta, \rho\}$*, and* μ_t *is linear in* p_t*. Furthermore, if the objective function from equation (2) is concave, then* $\lim_{T\to\infty} \lambda_t = \lambda$ *and* $\lim_{T\to\infty} \bar{\lambda}_t = \bar{\lambda}$ *for any fixed t. Finally,* $\lim_{T\to\infty} \mu_t = \mu$ *for any fixed t if* p_t *and* ξ_t *are constant and* $-\eta > (1-\alpha)\delta\rho\left[(\zeta-\eta)\left(1+\rho\bar{\lambda}_{t+1}\right)-\rho\zeta\right]$.

Proof. See Appendix F.2. That appendix also provides an explicit condition that guarantees concavity. □

Steady state. Over a period of time when strategies are well approximated by the limiting values λ and μ, consumption converges to a steady state.

Lemma 1. *Suppose that strategies in all periods take the form* $x_t^*(s_t, \gamma, p_t) = \lambda s_t + \mu$*, where* λ *and* μ *are constant. If* $\rho(1+\lambda) < 1$*, both* x_t^* *and* s_t *converge monotonically over time to steady-state values* x_{ss} *and* s_{ss}.

Proof. See Appendix F.3. □

If consumption has reached a steady state, we can use the Euler equation to characterize its level in closed form.

Proposition 3. *Suppose that* p_t *and* ξ_t *are constant and that consumption and habit stock are in steady state with* $s_t = s_{ss}$*,* $x_t = x_{ss}$*, and* $x_{ss} = \rho(s_{ss} + x_{ss})$*. Then consumption can be written as*

$$x_{ss} = \frac{\kappa - (1-(1-\alpha)\delta\rho)p + (1-\alpha)\delta\rho\left[(\zeta-\eta)m_{ss} - \left(1+\bar{\lambda}\right)\tilde{\gamma}\right] + \gamma}{-\eta - (1-\alpha)\delta\rho(\zeta-\eta) - \zeta\frac{\rho-(1-\alpha)\delta\rho^2}{1-\rho}}, \tag{21}$$

where $\kappa := (1-\alpha)\delta\rho(\phi-\xi)+\xi$ *and* $m_{ss} := \tilde{x}_{t+1} - x_{ss}$ *is steady-state misprediction.*

Proof. See Appendix F.4. □

The parameter restrictions required for Proposition 2 and Lemma 1 (including concavity) essentially amount to requiring that perceived and actual habit formation are not too strong. We have confirmed that these restrictions hold at the parameter estimates presented in Table 4.

E.2 Modeling the Experiment

We need additional notation to map the experiment's treatments and data into the model and estimation. We define x_{it} to be participant i's daily average FITSBY screen time during period t, $\tilde{x}_{it}$ to be participant i's predicted screen time elicited on a survey, and $m_{it} = x_{it} - \tilde{x}_{it}$ to be the difference between the two. The Bonus and Bonus Control groups are denoted $g \in \{B, BC\}$, the Limit and Limit Control groups are $g \in \{L, LC\}$, and the intersection of Bonus Control and Limit Control is $g = C$. We define $\bar{y} := \mathbb{E}_i y_i$ as the expectatation over participants of variable y, and $y^g := \mathbb{E}_{i \in g} y_i$ as the expectation over group g. $\tau_t^g := x_t^g - x_t^{gC}$ and $\tilde{\tau}_t^g := \tilde{x}_t^g - \tilde{x}_t^{gC}$ are the actual and predicted average treatment effects.

We model the Screen Time Bonus as a price $p^B = \$2.50$ per hour in period 3 plus a fixed payment $F_i^B = \$50 \times \mathrm{ceil}(x_{i1} \frac{\text{hours}}{\text{day}})$, where $\mathrm{ceil}(\cdot)$ rounds up to the nearest integer, giving participant i's Bonus Benchmark. In this appendix, we generalize the primary model from Section 6 by modeling the limit as an intervention that eliminates share ω of temptation.

We define v_i^B as the valuation of the bonus elicited on survey 2, and we define v_i^L as the valuation of access to the limit functionality elicited on survey 3. We assume that on survey t, consumers are aware of period t projection bias when predicting period t consumption and are projection biased when determining their bonus and limit valuations. This assumption means that misprediction of period-t consumption is driven only by naivete about temptation, and that bonus and limit valuations are driven only by perceived temptation, not by an additional desire to offset projection bias. We acknowledge that alternative assumptions could be made.

E.3 Estimating Equations

Using the theoretical results from Appendix E.1, we can now derive equations that characterize how a consumer from our unrestricted model would behave in our experiment. These equations parallel the equation in Section 6.2, with additional terms that account for perceived habit formation. We assume that the discount factor is $\delta = 0.997$ per three-week period, consistent with a five percent annual discount rate. We estimate the remaining parameters in stages, as described below. Appendix G presents formal derivations and additional details.

Habit Formation

We first estimate λ and ρ from the decay of the bonus treatment effects. Even though λ is not a structural parameter, it is easily identified and useful in estimating the other parameters. Using the habit stock evolution formula and the linearity result in equation (19), we can write the period 4 bonus effect as the result of decayed effects from periods 2 and 3: $\tau_4^B = \lambda\left(\rho\tau_3^B + \rho^2\tau_2^B\right)$. Similarly, the period 5 effect results from the cumulative decayed effects from periods 2–4: $\tau_5^B = \lambda\left(\rho\tau_4^B + \rho^2\tau_3^B + \rho^3\tau_2^B\right)$. Rearranging gives a system of two equations for λ and ρ:

$$\lambda = \frac{\tau_4^B}{\rho \tau_3^B + \rho^2 \tau_2^B} \tag{22}$$

$$\rho = \frac{\tau_5^B}{\tau_4^B (1+\lambda)}. \tag{23}$$

This non-linear system has two solutions when $\tau_2^B \neq 0$, but in our data there is only one solution that satisfies the requirement that $\rho \geq 0$.

For estimation, we assume $\tilde{\lambda} = \lambda$. This is reasonable because Figure 7 shows that participants predicted the time path of bonus effects with reasonable accuracy, so calibrating equations (22) and (23) with predicted τ_t^B would not change the estimates much. To the extent that predictions differ from actual behavior, we prefer to err on the side of using actual behavior instead of beliefs to estimate the model.

Perceived Habit Formation, Price Response, and Habit Stock Effect on Marginal Utility

After estimating λ and ρ, we estimate α, η, and ζ from the magnitude and decay of the bonus treatment effects. For each of periods 2, 3, and 4, we difference the Euler equations for the Bonus and Bonus Control groups and rearrange, giving a system of three equations for $(1-\alpha)$, η, and ζ:

$$(1-\alpha) = \frac{\eta \tau_2^B}{\delta \rho \left[-p^B + (\eta - \zeta)\tilde{\tau}_3^B + \zeta \rho \tau_2^B\right]}. \tag{24}$$

$$\eta = \frac{p^B - \zeta \rho \tau_2^B + (1-\alpha)\delta \rho^2 \zeta (1-\tilde{\lambda}) \left(\rho \tau_2^B + \tau_3^B\right)}{\tau_3^B - (1-\alpha)\delta \rho^2 \tilde{\lambda} \left(\rho \tau_2^B + \tau_3^B\right)} \tag{25}$$

$$\zeta = \frac{-\eta \tau_4^B + (1-\alpha)\delta \rho^2 \eta \tilde{\lambda} \left(\rho^2 \tau_2^B + \rho \tau_3^B + \tau_4^B\right)}{\rho \tau_3^B + \rho^2 \tau_2^B - (1-\alpha)\delta \rho^2 \left(1-\tilde{\lambda}\right) \left(\rho^2 \tau_2^B + \rho \tau_3^B + \tau_4^B\right)}. \tag{26}$$

The first equation shows that as the anticipatory demand response in period 2 grows compared to the predicted demand response in period 3 (making $\tau_2^B / \tilde{\tau}_3^B$ larger), we infer more perceived habit formation (smaller α).

Naivete about Temptation

Next, we estimate naivete about temptation $\gamma - \tilde{\gamma}$ using the Control group's difference between perceived and actual consumption. To solve for $\gamma - \tilde{\gamma}$, we difference the actual versus perceived Euler equations for group C, giving

$$\gamma - \tilde{\gamma} = m_t^C \cdot \left[-\eta + (1-\alpha)\delta \rho^2 \left((\eta - \zeta)\tilde{\lambda} + \zeta\right)\right]. \tag{27}$$

Temptation

We estimate temptation γ using three different strategies: the limit treatment effect and valuations of the bonus and limit. Each strategy delivers an equation that we combine with equation (27) to form a system of two equations for γ and $\tilde{\gamma}$.

Limit effect. Recall that we model the limit as an intervention that eliminates share ω of temptation, starting in period 2. Thus, we can identify γ using an assumed ω plus the effect of the limit on consumption. To solve for γ, we difference the Euler equations for periods 2 versus 3 for the Limit group compared to Limit Control and rearrange, giving

$$\gamma = \eta \tau_2^L/\omega - (1-\alpha)\delta\rho \left[(\eta-\zeta)\tilde{\tau}_3^L/\omega + \zeta\rho\tau_2^L/\omega - \tilde{\gamma} - \tilde{\gamma}\tilde{\lambda}\right]. \quad (28)$$

Our primary estimates in Section 6 use this equation, after setting $\omega = 1$ and $\alpha = 1$.

Bonus valuation. Since the bonus is like a commitment device that reduces future use, people with perceived self-control problems will place higher value on the bonus. We can estimate perceived temptation $\tilde{\gamma}$ from participants' valuations. Our derivation follows Allcott, Kim, Taubinsky, and Zinman (2021), and the approach also follows Acland and Levy (2012), Augenblick and Rabin (2019), Chaloupka, Levy, and White (2019), and Carrera et al. (2021).

Let $V_t(\tilde{s}_t, \cdot)$ be the period t continuation value function conditional on $\tilde{s}_t$, according to predicted consumption and preferences before period t. This reflects preferences of a consumer filling out the multiple price list on a survey before period t. Since utility is quasilinear in money, $V_t(s_t, \cdot)$ is in units of period t dollars.

The effect of a period 3 price increase from 0 to p_3^B on the period 3 continuation value is

$$\Delta V_3\left(p^B\right) := V_3\left(\tilde{s}_3, p_3 = p_3^B\right) - V_3\left(\tilde{s}_3, p_3 = 0\right) = -p_3^B \cdot \frac{1}{2}\left(\tilde{x}_3(p_3^B) + \tilde{x}_3(0)\right) - \tilde{\gamma} \cdot \left(\tilde{x}_3(p_3^B) - \tilde{x}_3(0)\right), \quad (29)$$

where $\tilde{x}_3(p_3) = \tilde{x}_3^*(\tilde{s}_3, \tilde{\gamma}, \boldsymbol{p}_3)$ is shorthand for predicted period 3 consumption as a function of period 3 price. Figure 9 illustrates. The trapezoid $ABCD$ is $p_3^B \cdot \frac{1}{2}\left(\tilde{x}_3(p_3^B) + \tilde{x}_3(0)\right)$: the survey taker's prediction of the consumer surplus loss from the price increase from the period 3 self's perspective. The parallelogram $BCEF$ is $-\tilde{\gamma} \cdot \left(\tilde{x}_3(p_3^B) - \tilde{x}_3(0)\right)$: the predicted additional temptation reduction benefit from the survey taker's perspective.

The Screen Time Bonus combines a price change with a fixed payment of F^B. Thus, the model predicts that people filling out the bonus MPL would be indifferent between the bonus and a fixed payment of $v^B = F^B + \Delta V_3(p^B)$. Taking the expectation over participants to allow mean-zero survey noise, substituting $\tilde{\tau}_3^B := \mathbb{E}_i\left[\tilde{x}_{i3}(p_3^B) - \tilde{x}_{i3}(0)\right]$ and $\bar{\tilde{x}}_3^{B+BC} := \mathbb{E}_i\left[\frac{1}{2}\left(\tilde{x}_{i3}(p_3^B) + \tilde{x}_{i3}(0)\right)\right]$, and rearranging gives perceived temptation:

$$\tilde{\gamma} = \frac{\bar{v}^B - \bar{F}^B + p_3^B \bar{\tilde{x}}_3^{B+BC}}{-\tilde{\tau}_3^B}. \quad (30)$$

The model predicts that if consumers perceive themselves to be time consistent ($\tilde{\gamma} = 0$), the average bonus valuation would equal the average valuation from the period 3 self's perspective, $\bar{F}^B - p_3^B \bar{x}_3^{B+BC}$. We refer to the difference between the observed average valuation and the modeled time-consistent valuation (the numerator of equation (30)) as "behavior change premium." We infer more perceived temptation $\tilde{\gamma}$ from a larger behavior change premium.

Limit valuation. People who perceive future temptation value the limit, as they perceive that it eliminates share ω of temptation. We can estimate perceived temptation $\tilde{\gamma}$ using an assumed ω plus the valuation the limit functionality. We solve for the modeled valuation similarly to how we solved for the bonus valuation above.

The effect of a period 3 temptation reduction from $\tilde{\gamma}$ to $(1-\omega)\tilde{\gamma}$ on the period 3 continuation value is

$$v^L = V_3(s_3, \tilde{\gamma}_3 = (1-\omega)\tilde{\gamma}) - V_3(s_3, \tilde{\gamma}_3 = \tilde{\gamma}) = \tilde{\gamma} \cdot (x_3^*(\tilde{\gamma}) - x_3^*((1-\omega)\tilde{\gamma})) \cdot \frac{2-\omega}{2}, \tag{31}$$

where $x_3^*(\tilde{\gamma}_3)$ is now shorthand for predicted period 3 consumption as a function of predicted period 3 temptation. Figure 9 illustrates. With $\omega = 1$, the limit valuation is the deadweight loss reduction CEG from the survey taker's perspective from consuming the desired amount ($x_3^*(0)$, point G) instead of the predicted amount ($x_3^*(\tilde{\gamma})$, point C). The height of this triangle is $\tilde{\gamma}$ and the width is $x_3^*(\tilde{\gamma}) - x_3^*(0)$, and thus the area is $\tilde{\gamma} \cdot (x_3^*(\tilde{\gamma}) - x_3^*(0)) \cdot \frac{1}{2}$. With $\omega < 1$, the valuation v^L equals the deadweight loss reduction trapezoid starting to the right of point G and bounded by segment CE.

Taking the expectation over participants, substituting $\bar{\tau}_3^L := \mathbb{E}_i[x_3^*((1-\omega)\tilde{\gamma}) - x_3^*(\tilde{\gamma})]$, and rearranging gives perceived temptation:

$$\tilde{\gamma} = \frac{\bar{v}^L}{-\bar{\tau}_3^L(2-\omega)/2}. \tag{32}$$

We infer more perceived temptation $\tilde{\gamma}$ from higher valuation $\bar{v}^L$.

Intercept

Finally, we back out a heterogeneous intercept κ_i that explains observed consumption heterogeneity. Our data do not allow us to separately identify ϕ (the direct effect of habit stock on utility) from ξ (the marginal utility shifter), so κ_i includes both of these structural parameters. We assume that participant i's observed baseline consumption x_{i1} is in a steady state characterized by equation (21). Rearranging that equation gives

$$\kappa_i := (1-\alpha)\delta\rho(\phi - \xi_i) + \xi_i = (1-(1-\alpha)\delta\rho)\,p - (1-\alpha)\delta\rho\left[(\zeta-\eta)m_{ss} - \left(1+\tilde{\lambda}\right)\tilde{\gamma}\right]$$
$$-\gamma + x_{i1}\left[-\eta - (1-\alpha)\delta\rho(\zeta-\eta) - \zeta\frac{\rho-(1-\alpha)\delta\rho^2}{1-\rho}\right]. \tag{33}$$

E.4 Empirical Moments and Estimation Details

Appendix Table A7 presents the full set of moments and fixed parameter values that are inputs to our unrestricted model and alternative specifications. In light of the discussion in Section 5.3, we omit the first half of period 2 when we estimate the anticipatory bonus effect τ_2^B.[22] The average of predicted use with and without the bonus $\tilde{x}_3^{B,BC}$ and the predicted contemporaneous bonus effect $\tilde{\tau}_3^B$ are the predictions before the bonus MPL on survey 2, as displayed in Figure 7. Because we do not have an explicit elicitation of the predicted limit effect, we use the actual limit effect τ_3^L to proxy for the predicted limit effect $\tilde{\tau}_3^L$.[23] Since Figure 6 shows that the average prediction error for period t consumption is similar when elicited on survey t versus survey $t-1$, we let observed Control group misprediction m^C proxy for steady-state misprediction m_{ss}.

We winsorize the anticipatory bonus effect at $\tau_2^B \leq 0$, which affects 15 percent of draws. We also drop the 0.32 percent of bootstrap draws in which the denominator of steady-state consumption in equation (15) is not positive.

[22]Appendix Table A8 presents parameter estimates when we use all of period 2 to estimate τ_2^B. The estimated ρ is larger, as expected, but the other parameter estimates are very similar.

[23]The average difference in predicted FITSBY use between Limit and Limit Control on survey 3 is $\tilde{\tau}_3^L \approx -10.5$ minutes per day, much smaller than the actual limit effect of $\tau_3^L \approx -22.3$ minutes per day. In the limit effect strategy in equation (28), $\tilde{\tau}_3^L$ makes little difference because it is multiplied by $(1-\alpha)$, which is small. However, in the limit valuation strategy in equation (32), $\tilde{\gamma}$ is inversely proportional to $\tilde{\tau}_3^L$, so a much smaller $\tilde{\tau}_3^L$ would make the estimated $\tilde{\gamma}$ much larger.

Table A7: **Empirical Moments and Additional Parameters**

Parameter	Description	(1) Point estimate	(2) Confidence interval
δ	Three-week discount factor (unitless)	0.997	
τ_2^B	Anticipatory bonus effect (minutes/day)	-1.96	$[-7.40, 0]$
τ_3^B	Contemporaneous bonus effect (minutes/day)	-55.9	[-61.7, -50.3]
τ_4^B	Long-term bonus effect (minutes/day)	-19.2	[-24.7, -13.7]
τ_5^B	Long-term bonus effect (minutes/day)	-12.3	[-18.1, -6.54]
τ_2^L	Limit effect (minutes/day)	-24.3	[-28.1, -20.4]
m^C, m_{ss}	Control group misprediction (minutes/day)	6.13	$[4.52, 7.72]$
$\tilde{x}_3^{B+BC}$	Predicted use with/without bonus (minutes/day)	122	$[114, 130]$
$\tilde{\tau}_3^B$	Predicted bonus effect (minutes/day)	-45.0	$[-50.0, -40.1]$
$\tilde{\tau}_3^L$	Predicted limit effect (minutes/day)	-22.3	$[-27.3, -17.3]$
ω	Temptation reduction from limit	1	
$\bar{v}^B$	Average bonus valuation ($/day)	3.20	$[3.12, 3.29]$
$\bar{v}^L$	Average limit valuation ($/day)	0.210	[0.184, 0.237]
p^B	Bonus price ($/hour)	2.5	
$\bar{F}^B$	Average bonus fixed payment ($/day)	7.03	$[6.96, 7.09]$
$\bar{x}_1$	Average baseline use (minutes/day)	153	$[149, 157]$

Notes: This table presents point estimates and bootstrapped 95 percent confidence intervals for the empirical moments used for estimation. We winsorize at $\tau_2^B \leq 0$, and we drop the 0.32 percent of draws in which the denominator of steady-state consumption in equation (15) is not positive.

Table A8: **Primary Parameter Estimates Using τ_2^B for All of Period 2**

Parameter	Description (units)	(1) Unrestricted model ($\alpha = \tilde{\alpha}$)
λ	Habit stock effect on consumption (unitless)	1.08
		$[0.565, 3.09]$
ρ	Habit formation (unitless)	0.308
		$[0.113, 0.507]$
α	Projection bias (unitless)	0.725
		$[0.427, 0.969]$
η	Price coefficient (\$-day/hour^2)	-2.85
		$[-3.15, -2.61]$
ζ	Habit stock effect on marginal utility (\$-day/hour^2)	2.91
		$[1.49, 8.45]$
$\gamma - \tilde{\gamma}$	Naivete about temptation (\$/hour)	0.283
		$[0.208, 0.359]$
γ	Temptation (\$/hour)	1.16
		$[0.938, 1.40]$
$\bar{\kappa}$	Average intercept (\$/hour)	-1.95
		$[-3.40, -0.574]$

Notes: This table presents point estimates and bootstrapped 95 percent confidence intervals from the estimation strategy described in Section E.3. We winsorize at $\tau_2^B \leq 0$, and we drop the 0.32 percent of draws in which the denominator of steady-state consumption in equation (15) is not positive. Temptation γ is from the limit effect strategy, using equation (28). This parallels column 2 of Table 4, except using all of period 2 (instead of only the second half of period 2) to estimate the anticipatory bonus effect τ_2^B.

E.5 Alternative Temptation Estimates

Appendix Table A9 presents alternative estimates of temptation γ in the restricted and unrestricted models. After repeating the primary limit effect estimate, the table reports the bonus valuation estimate. Before the bonus MPL on survey 2, the average participant predicted that they would use FITSBY 2.5 and 1.6 hours per day without and with the bonus, respectively. Thus, the average survey taker would have predicted that the price increase would cause a consumer surplus loss from their period 3 self's perspective of $p_3^B \bar{x}_3 \approx \$2.50 \times \frac{1}{2}(2.5 + 1.6) \approx \$5.09$ per day of period 3. This is the trapezoid *ABCD* on Figure 9. The average bonus fixed payment was $\bar{F}^B \approx \$7.03$ per day. Thus, if the average participant perceived herself to be time consistent, she would have been indifferent between the bonus and a certain payment of $\$7.03 - \$5.09 \approx \$1.94$ per day.

In reality, the average participant was indifferent between the bonus and a certain payment of \$64, or $\bar{v}^B \approx \$64/20 \approx \$3.20$ per day over the 20-day period. This excess valuation implies a behavior change

premium of $\$3.20 - \$1.94 \approx \$1.26$ per day. This is the parallelogram *BCEF* on Figure 9: the additional temptation reduction benefit that the period 2 survey taker perceives from the reduced FITSBY use caused by the bonus. Rearranging this logic into equation (30) gives perceived temptation $\hat{\tilde{\gamma}} \approx 1.34$ \$/hour. Using the estimated naivete of $\widehat{\gamma - \tilde{\gamma}} \approx 0.274$ gives $\hat{\gamma} \approx 1.61$ for the bonus valuation strategy in column 1.

The average Limit group participant was indifferent between access to the limit functionality for period 3 and a certain payment of \$4.20, or $\tilde{v}^L \approx \$4.20/20 \approx \$0.210$ per day over the 20-day period. This is the triangle on Figure 9: the perceived deadweight loss reduction from the reduced FITSBY use caused by the limit. Inserting this into equation (32) with $\omega = 1$ gives perceived temptation $\hat{\tilde{\gamma}} = \frac{\tilde{v}^L}{-\tau_3^L/2} \approx \frac{0.210}{(-(-22.3)/60)/2} \approx$ 1.13 \$/hour. Using $\widehat{\gamma - \tilde{\gamma}} \approx 0.274$ gives $\hat{\gamma} \approx 1.41$ for the limit valuation strategy in column 1.

So far, we have modeled FITSBY screen time on other devices as part of an outside option that is not affected by self-control problems. In Appendix G.5, we generalize the model to include multiple temptation goods. As discussed in Section 5.5, self-reports suggest that the limit increased FITSBY use on other devices by 4.2 minutes per day, while the bonus reduced FITSBY use on other devices by 8.1 minutes per day. We use these additional moments to identify the multiple-good model.

The next three rows in Appendix Table A9 present estimates from the multiple-good model. The limit effect estimate increases to $\hat{\gamma} \approx 1.31$ \$/hour, because in the multiple-good model, more temptation is needed to explain the observed limits when consumers setting the limits think they'll evade the limits through substitution to other devices. The bonus valuation estimate decreases to $\hat{\gamma} \approx 1.44$ \$/hour, because in the multiple-good model, less temptation is needed to explain the observed bonus valuation when consumers think the bonus will also reduce FITSBY use on other devices. The limit valuation estimate increases to to $\hat{\gamma} \approx 2.09$ \$/hour, because in the multiple-good model, more temptation is needed to explain the observed limit valuation when consumers think the limit will also increase FITSBY use on other devices.

Next, we return to the single-good model and consider an alternative specification where we estimate ω from differences in self-reported *ideal use change* between the Limit and Limit Control groups. Intuitively, if the Limit group reports on survey 3 that looking back over period 2, they ideally would not have further reduced their screen time, this suggests that the limit functionality fully eliminated temptation ($\omega = 1$). Extending this intuition, we estimate ω as the share of the Limit Control group's *ideal use change* that is eliminated in the Limit treatment group. If d_2^g is group g's average *ideal use change* reported on survey 3 retrospectively about period 2, this is:

$$\omega = \frac{d_2^L - d_2^{LC}}{-d_2^{LC}}. \tag{34}$$

In the data, the Limit and Limit Control groups report that they ideally would have changed use by -9.5 and -15 percent, respectively. This gives $\hat{\omega} \approx \frac{-0.095-(-0.15)}{-(-0.15)} \approx 0.385$.

If we assume that the limit only eliminates share $\omega < 1$ of temptation, the limit effect strategy will deliver larger γ, because we infer that the true effect of temptation on consumption is larger. By contrast, the limit valuation strategy will deliver smaller γ, because a smaller γ is needed to explain a given valuation $\tilde{v}^L$ when temptation has a larger effect on consumption. Appendix Table A9 shows that in the restricted model

($\alpha = 1$), the limit effect $\hat{\gamma}$ increases from 1.09 to 2.82, while the limit valuation strategy $\hat{\gamma}$ decreases from 1.41 to 0.975.

Finally, we extend the limit effect strategy to allow for individual-specific heterogeneity in γ. To do this, we exploit the facts that we observe each participant's period 2 *limit tightness* H_{i2} and that tightness is closely related to the limit treatment effect. We estimate heterogeneous period 2 and 3 limit effects as a function of period 2 *limit tighness* by adding an interaction term $\tau^{HL} H_{i2} L_i$ to the treatment effect estimation in equation (4); see Appendix Table A10.[24] For each participant, we insert the fitted limit effect $\hat{\tau}^L_{it} = \hat{\tau}^L_t + \hat{\tau}^{HL} H_{i2}$ into equation (28) to infer γ_i. The final row of Appendix Table A9 shows that although this allows substantial heterogeneity, the average temptation $\bar{\gamma}$ is essentially the same as the homogeneous γ from the limit effect strategy, as one would expect.

These alternative approaches imply temptation γ is between about \$1 and \$3 per hour. Our primary strategy (the limit effect) is relatively conservative.

[24] H_i is missing for the Limit Control group, so we are not able to include the main effect of H_{i2} in this regression. In theory, this could generate omitted variable bias if period 2 or 3 control group consumption varies with the tightness that they would have set. Appendix Table A10 shows that H_{i2} is associated with the Limit group's consumption in the second half of period 1 (before the limit functionality was turned on). However, the association is small compared to the association in periods 2 and 3, which suggests that the potential omitted variables bias is relatively small.

Table A9: **Alternative Temptation Parameter Estimates**

Parameter	Description (units)	(1) Restricted model ($\tau_2^B = 0$, $\alpha = 1$)	(2) Unrestricted model ($\alpha = \hat{\alpha}$)
γ	Temptation ($/hour)		
	Limit effect (primary)	1.09	1.11
		[0.884, 1.30]	[0.903, 1.33]
	Bonus valuation	1.61	1.62
		[1.29, 1.94]	[1.29, 1.94]
	Limit valuation	1.41	1.41
		[1.19, 1.75]	[1.19, 1.76]
	Limit effect, multiple-good model	1.31	
		[1.01, 1.71]	
	Bonus valuation, multiple-good model	1.44	1.45
		[1.16, 1.73]	[1.17, 1.74]
	Limit valuation, multiple-good model	2.09	2.09
		[1.33, 7.10]	[1.33, 7.10]
	Limit effect, $\omega = \hat{\omega}$	2.82	2.92
		[2.11, 3.92]	[2.22, 4.16]
	Limit valuation, $\omega = \hat{\omega}$	0.975	0.979
		[0.826, 1.19]	[0.833, 1.20]
$\tilde{\gamma}$	Average temptation ($/hour)	1.08	1.10
	Heterogeneous limit effect	[0.873, 1.29]	[0.889, 1.31]

Notes: This table presents point estimates and bootstrapped 95 percent confidence intervals for alternative estimates of temptation γ. Each row reflects estimates from a different specification. γ for the limit effect, bonus valuation, and limit valuation strategies is from equations (28), (30), and (32), respectively, combined with naivete $\gamma - \tilde{\gamma}$ from equation (27). γ for the multiple-good model is from equations (182), (187), and (190) in Appendix G.5; we do not have a limit effect estimate for the unrestricted multiple-good model. $\hat{\omega}$ is from equation (34).

Table A10: **Heterogeneity in Limit Effect by Limit Tightness**

	2nd half of period 1 FITSBY use	Period 2 FITSBY use	Period 3 FITSBY use
	(1)	(2)	(3)
Bonus treatment	−4.702	−3.228	−54.384
	(2.001)	(2.154)	(2.835)
Limit treatment	−5.281	0.447	−1.248
	(2.143)	(2.308)	(3.041)
Limit treatment × period 2 limit tightness	0.114	−0.551	−0.469
	(0.027)	(0.029)	(0.039)
1st half of period 1 FITSBY use	0.845		
	(0.014)		
Period 1 FITSBY use		0.894	0.795
		(0.015)	(0.020)
Observations	1,933	1,930	1,931
R^2	0.849	0.795	0.665

Notes: This table presents the effects of bonus and limit treatments on FITSBY use in periods 1, 2, and 3 using equation (4), including an additional interaction between the Limit group indicator and period 2 *limit tightness*. *Limit tightness* is the amount by which a user's limits would have hypothetically reduced overall screen time if applied to their baseline use without snoozes; see equation (5). FITSBY use refers to screen time on Facebook, Instagram, Twitter, Snapchat, browser, and YouTube.

E.6 Model Estimates with Sample Weights

Table A11: **Demographics in Weighted Sample**

	(1) Analysis sample	(2) Balanced sample	(3) U.S. adults
Income ($000s)	40.8	42.1	43.0
College	0.67	0.55	0.30
Male	0.39	0.42	0.49
White	0.72	0.72	0.74
Age	33.7	38.7	47.6
Period 1 phone use (minutes/day)	333.0	339.3	.
Period 1 FITSBY use (minutes/day)	152.8	155.4	.

Notes: Column 1 presents average demographics for our analysis sample, column 2 presents average demographics for our weighted sample, and column 3 presents average demographics of American adults using data from the 2018 American Community Survey. The sample weights are initially calculated to make the sample nationally representative on these five demographics but are then winsorized at $[1/3, 3]$ to reduce precision loss.

Table A12: **Empirical Moments and Additional Parameters in Weighted Sample**

Parameter	Description	(1) Point estimate	(2) Confidence interval
δ	Three-week discount factor (unitless)	0.997	
τ_2^B	Anticipatory bonus effect (minutes/day)	-4.41	$[-12.8, 0]$
τ_3^B	Contemporaneous bonus effect (minutes/day)	-58.5	[-67.3, -50.3]
τ_4^B	Long-term bonus effect (minutes/day)	-25.1	[-34.7, -15.7]
τ_5^B	Long-term bonus effect (minutes/day)	-16.4	[-26.5, -7.93]
τ_2^L	Limit effect (minutes/day)	-23.3	[-29.6, -16.7]
m^C	Control group misprediction (minutes/day)	4.96	$[3.03, 7.11]$
$\tilde{\bar{x}}_3^{B+BC}$	Predicted use with/without bonus (minutes/day)	127	$[114, 140]$
$\tilde{\tau}_3^B$	Predicted bonus effect (minutes/day)	-49.4	$[-56.6, -41.9]$
$\tilde{\tau}_3^L$	Predicted limit effect (minutes/day)	-20.7	$[-27.9, -12.7]$
ω	Temptation reduction from limit	1	
$\bar{v}^B$	Average bonus valuation ($/day)	3.29	$[3.15, 3.44]$
$\bar{v}^L$	Average limit valuation ($/day)	0.271	[0.229, 0.315]
p^B	Bonus price ($/hour)	2.5	
$\bar{F}^B$	Average bonus fixed payment ($/day)	6.84	$[6.72, 6.96]$
$\bar{x}_1$	Average baseline use (minutes/day)	156	$[149, 164]$

Notes: This table presents point estimates and bootstrapped 95 percent confidence intervals for the empirical moments used for estimation. We winsorize at $\tau_2^B \leq 0$, and we drop the 0.32 percent of draws in which the denominator of steady-state consumption in equation (15) is not positive. This parallels Table 3, except using the weighted sample. The sample weights are initially calculated to make the sample nationally representative on the five demographics in Appendix Table A11 but are then winsorized at $[1/3, 3]$ to reduce precision loss.

Table A13: **Model Parameter Estimates in Weighted Sample**

Parameter	Description (units)	Restricted model ($\tau_2^B = 0$, $\alpha = 1$)
λ	Habit stock effect on consumption (unitless)	1.93
		[0.757, 3.89]
ρ	Habit formation (unitless)	0.223
		[0.122, 0.469]
α	Projection bias (unitless)	1
η	Price coefficient (\$-day/hour^2)	-2.57
		[-2.98, -2.23]
ζ	Habit stock effect on marginal utility (\$-day/hour^2)	4.95
		[2.07, 9.96]
$\gamma - \tilde{\gamma}$	Naivete about temptation (\$/hour)	0.212
		[0.130, 0.307]
γ	Temptation (\$/hour)	0.998
		[0.709, 1.30]
$\bar{\kappa}$	Average intercept (\$/hour)	-1.99
		[-3.52, -0.422]

Notes: This table presents point estimates and bootstrapped 95 percent confidence intervals from the estimation strategy described in Section E.3. We winsorize at $\tau_2^B \leq 0$, and we drop the 0.32 percent of draws in which the denominator of steady-state consumption in equation (15) is not positive. Temptation γ is from the limit effect strategy, using equation (28). This parallels Table 4, except using the weighted sample. The sample weights are initially calculated to make the sample nationally representative on the five demographics in Appendix Table A11 but are then winsorized at [1/3, 3] to reduce precision loss.

F Proofs of Propositions in Appendix E.1

Given naivete about projection bias, the predicted continuation value function given predicted consumption and habit stock is

$$V_{t+1}(\tilde{s}_{t+1}) = \sum_{r=t+1}^{T} \delta^{r-t} u_r(\tilde{x}_r^*(\tilde{s}_r, \tilde{\gamma}, p_r); \tilde{s}_r, p_r). \tag{35}$$

The consumer's predicted objective function in future period t can thus be written as

$$\tilde{U}_t(x_t; \tilde{s}_t) = u_t(x_t; \tilde{s}_t, p_t) + \tilde{\gamma} x_t + \delta V_{t+1}(\tilde{s}_{t+1}), \tag{36}$$

and the consumer's actual period t objective function from equation (2) can be written as

$$U_t(x_t;s_t) = u_t(x_t;s_t,p_t) + \gamma x_t + \begin{array}{c} \alpha \sum_{r=t+1}^{T} \delta^{r-t} u_r\left(\tilde{x}_r^*(s_t,\tilde{\gamma},\boldsymbol{p}_r);s_t,p_r\right) \\ +(1-\alpha)\delta V_{t+1}(\tilde{s}_{t+1}) \end{array}. \tag{37}$$

Recall that we defined $u_t := u_t\left(x_t^*;s_t,p_t\right)$, $\tilde{x}_r := \tilde{x}_r^*\left(\tilde{s}_r,\tilde{\gamma},\boldsymbol{p}_r\right)$, and $\tilde{u}_r := u_r\left(\tilde{x}_r;\tilde{s}_r,p_r\right)$.

F.1 Proof of Proposition 1: Euler Equation

In this section, we derive the Euler equation (equation (18)), proving Proposition 1.

Proof. The time t first-order condition from maximizing utility (equation (37)) is

$$\frac{\partial u_t}{\partial x_t} + \gamma = -(1-\alpha)\delta \frac{d\tilde{s}_{t+1}}{dx_t}\frac{dV_{t+1}(\tilde{s}_{t+1})}{d\tilde{s}_{t+1}} \tag{38}$$

$$= -(1-\alpha)\delta \frac{d\tilde{s}_{t+1}}{dx_{t+1}}\left[\frac{\partial\tilde{u}_{t+1}}{\partial\tilde{x}_{t+1}}\frac{\partial\tilde{x}_{t+1}}{\partial\tilde{s}_{t+1}} + \frac{\partial\tilde{u}_{t+1}}{\partial\tilde{s}_{t+1}}\right] - (1-\alpha)\delta^2\frac{d\tilde{s}_{t+2}}{dx_t}\frac{dV_{t+2}(\tilde{s}_{t+2})}{d\tilde{s}_{t+2}} \tag{39}$$

$$= -(1-\alpha)\delta\rho\left[\frac{\partial\tilde{u}_{t+1}}{\partial\tilde{x}_{t+1}}\frac{\partial\tilde{x}_{t+1}}{\partial\tilde{s}_{t+1}} + \frac{\partial\tilde{u}_{t+1}}{\partial\tilde{s}_{t+1}}\right] - (1-\alpha)(\delta\rho)^2\left(1+\frac{\partial\tilde{x}_{t+1}}{\partial\tilde{s}_{t+1}}\right)\frac{dV_{t+2}(\tilde{s}_{t+2})}{d\tilde{s}_{t+2}}, \tag{40}$$

where the third line uses the fact that the total derivative of predicted period $t+2$ habit stock with respect to period t consumption is

$$\frac{d\tilde{s}_{t+2}}{dx_t} = \frac{\partial\tilde{s}_{t+2}}{\partial\tilde{s}_{t+1}}\frac{\partial\tilde{s}_{t+1}}{\partial x_t} + \frac{\partial\tilde{s}_{t+2}}{\partial\tilde{x}_{t+1}}\frac{\partial\tilde{x}_{t+1}}{\partial\tilde{s}_{t+1}}\frac{\partial\tilde{s}_{t+1}}{\partial x_t}$$
$$= \rho^2\left(1+\frac{\partial\tilde{x}_{t+1}}{\partial\tilde{s}_{t+1}}\right) \tag{41}$$

The time t self predicts that the time $t+1$ self will maximize equation (36), setting x_{t+1} according to the following first-order condition:

$$0 = \frac{\partial\tilde{u}_{t+1}}{\partial\tilde{x}_{t+1}} + \tilde{\gamma} + \delta\frac{d\tilde{s}_{t+2}}{dx_{t+1}}\frac{dV_{t+2}(\tilde{s}_{t+2})}{d\tilde{s}_{t+2}} \tag{42}$$

$$= \frac{\partial\tilde{u}_{t+1}}{\partial\tilde{x}_{t+1}} + \tilde{\gamma} + \delta\rho\frac{dV_{t+2}(\tilde{s}_{t+2})}{d\tilde{s}_{t+2}} \tag{43}$$

Multiplying the predicted time $t+1$ first-order condition by $(1-\alpha)\delta\rho\left(1+\frac{\partial\tilde{x}_{t+1}}{\partial\tilde{s}_{t+1}}\right)$ gives

$$0 = (1-\alpha)\delta\rho\left(1+\frac{\partial\tilde{x}_{t+1}}{\partial\tilde{s}_{t+1}}\right)\left(\frac{\partial\tilde{u}_{t+1}}{\partial\tilde{x}_{t+1}}+\tilde{\gamma}\right) + (1-\alpha)(\delta\rho)^2\left(1+\frac{\partial\tilde{x}_{t+1}}{\partial\tilde{s}_{t+1}}\right)\frac{dV_{t+2}(\tilde{s}_{t+2})}{d\tilde{s}_{t+2}} \tag{44}$$

The last term is the same as the last term in the time t first-order condition. Adding this equation to the time t first-order condition yields

$$\frac{\partial u_t}{\partial x_t}+\gamma=(1-\alpha)\delta\rho\left(1+\frac{\partial\tilde{x}_{t+1}}{\partial\tilde{s}_{t+1}}\right)\left(\frac{\partial\tilde{u}_{t+1}}{\partial\tilde{x}_{t+1}}+\tilde{\gamma}\right)-(1-\alpha)\delta\rho\left[\frac{\partial\tilde{u}_{t+1}}{\partial\tilde{x}_{t+1}}\frac{\partial\tilde{x}_{t+1}}{\partial\tilde{s}_{t+1}}+\frac{\partial\tilde{u}_{t+1}}{\partial\tilde{s}_{t+1}}\right] \tag{45}$$

$$\frac{\partial u_t}{\partial x_t}+\gamma=(1-\alpha)\delta\rho\left[\frac{\partial\tilde{u}_{t+1}}{\partial\tilde{x}_{t+1}}+\tilde{\gamma}+\frac{\partial\tilde{x}_{t+1}}{\partial\tilde{s}_{t+1}}\tilde{\gamma}-\frac{\partial\tilde{u}_{t+1}}{\partial\tilde{s}_{t+1}}\right]. \tag{46}$$

We now derive the Euler equation with our quadratic functional form. The partial derivatives are

$$\frac{\partial u_t}{\partial x_t}=\eta x_t^*+\zeta s_t+\xi_t-p_t \tag{47}$$

$$\frac{\partial\tilde{u}_{t+1}}{\partial\tilde{x}_{t+1}}=\eta\tilde{x}_{t+1}+\zeta\tilde{s}_{t+1}+\xi_{t+1}-p_{t+1} \tag{48}$$

$$\tilde{\lambda}_{t+1}:=\frac{\partial\tilde{x}_{t+1}}{\partial\tilde{s}_{t+1}} \tag{49}$$

$$\frac{\partial\tilde{u}_{t+1}}{\partial\tilde{s}_{t+1}}=\zeta\tilde{x}_{t+1}+\phi. \tag{50}$$

Substituting these into equation (46) yields equation (18).

□

F.2 Proof of Proposition 2: Linear Policy Functions

In this section, we first show that the policy function is linear in habit stock. We then show that if the objective function is concave, λ converges to a constant far from the time horizon. We then show the conditions under which utility is concave. Finally, we show the condition required for μ to converge to a constant far from the time horizon. Our proof strategy follows Gruber and Köszegi (2001).

Lemma 2. *Suppose $u_t(x_t;s_t,p_t)$ is given by equation (3) and $(x_0^*,...,x_T^*)$ is a perception-perfect strategy profile. Then for any t,*

$$x_t^*(s_t,\gamma,\boldsymbol{p}_t)=\lambda_t s_t+\mu_t(\gamma) \tag{51}$$

$$\tilde{x}_t^*(s_t,\tilde{\gamma},\boldsymbol{p}_t)=\tilde{\lambda}_t s_t+\mu_t(\tilde{\gamma}) \tag{52}$$

where λ_t is a function of only $\{\eta,\zeta,\delta,\rho,\alpha\}$, $\tilde{\lambda}_t$ is a function of only $\{\eta,\zeta,\delta,\rho\}$, and μ_t is linear in p_t.

Proof. We prove by backwards induction. First, we show that the result holds for period T. Given our functional form, the period T first-order condition is

$$\eta x_T^* + \zeta s_T + \xi_T - p_T + \gamma = 0, \tag{53}$$

and thus

$$x_T^* = \frac{\zeta s_T + \xi_T - p_T + \gamma}{-\eta}. \tag{54}$$

Thus, x_T^* can be written as

$$x_T^* = \lambda_T s_T + \mu_T(\gamma), \tag{55}$$

with $\lambda_T = \frac{\zeta}{-\eta}$ and $\mu_T(\gamma) = \frac{\xi_T - p_T + \gamma}{-\eta}$.

Analogously, predicted consumption is

$$\tilde{x}_T = \frac{\zeta \tilde{s}_T + \xi_T - p_T + \tilde{\gamma}}{-\eta}, \tag{56}$$

so $\tilde{x}_T$ can be written as

$$\tilde{x}_T = \lambda_T \tilde{s}_T + \mu_T(\tilde{\gamma}), \tag{57}$$

with $\mu_T(\tilde{\gamma}) = \frac{\xi_T - p_T + \tilde{\gamma}}{-\eta}$. The function μ_T is linear in p_T.

Now, we use the Euler equation to show that if the result holds for $t+1$, it holds for t. The Euler equation is

$$\begin{aligned}\eta x_t^* + \zeta s_t + \xi_t - p_t + \gamma &= (1-\alpha)\delta\rho\left[\eta \tilde{x}_{t+1} + \zeta \tilde{s}_{t+1} + \xi_{t+1} - p_{t+1} + \left(1 + \frac{\partial \tilde{x}_{t+1}}{\partial \tilde{s}_{t+1}}\right)\tilde{\gamma} - \zeta \tilde{x}_{t+1} - \phi\right] \\ &= (1-\alpha)\delta\rho\left[(\eta - \zeta)\tilde{x}_{t+1} + \zeta \tilde{s}_{t+1} + \xi_{t+1} - p_{t+1} + \left(1 + \frac{\partial \tilde{x}_{t+1}}{\partial \tilde{s}_{t+1}}\right)\tilde{\gamma} - \phi\right]\end{aligned}$$

Substituting $\tilde{x}_{t+1} = \tilde{\lambda}_{t+1}\tilde{s}_{t+1} + \mu_{t+1}(\tilde{\gamma})$, $\tilde{s}_{t+1} = \rho\,(s_t + x_t^*)$, and $\tilde{\lambda}_{t+1} = \frac{\partial \tilde{x}_{t+1}}{\partial \tilde{s}_{t+1}}$ gives

$$\eta x_t^* + \zeta s_t + \xi_t - p_t + \gamma = (1-\alpha)\delta\rho\left[(\eta - \zeta)\left(\tilde{\lambda}_{t+1}\rho\,(x_t^* + s_t) + \mu_{t+1}(\tilde{\gamma})\right) + \zeta\rho(s_t + x_t^*) + \xi_{t+1} - p_{t+1} + \tilde{\gamma} + \tilde{\gamma}\tilde{\lambda}_{t+1} - \phi\right]. \tag{58}$$

Solving for x_t^* gives

$$x_t^* = \frac{s_t\left[\zeta - (1-\alpha)\delta\rho^2\left((\eta-\zeta)\tilde{\lambda}_{t+1} + \zeta\right)\right] + \xi_t - p_t + \gamma - (1-\alpha)\delta\rho\left[(\eta-\zeta)\mu_{t+1}(\tilde{\gamma}) + \xi_{t+1} - p_{t+1} + \tilde{\gamma} + \tilde{\gamma}\tilde{\lambda}_{t+1} - \phi\right]}{-\eta + (1-\alpha)\delta\rho^2\left((\eta-\zeta)\tilde{\lambda}_{t+1} + \zeta\right)}. \tag{59}$$

Thus, $x_t^* = \lambda_t s_t + \mu_t(\gamma)$, with

$$\lambda_t = \frac{\zeta - (1-\alpha)\delta\rho^2\left((\eta-\zeta)\bar{\lambda}_{t+1}+\zeta\right)}{-\eta+(1-\alpha)\delta\rho^2\left((\eta-\zeta)\bar{\lambda}_{t+1}+\zeta\right)}, \tag{60}$$

and

$$\mu_t(\gamma) = \frac{\xi_t - p_t + \gamma - (1-\alpha)\delta\rho\left[\xi_{t+1} - p_{t+1} + \tilde{\gamma} + \tilde{\gamma}\bar{\lambda}_{t+1} - \phi\right] + (1-\alpha)\delta\rho\,(\zeta-\eta)\,\mu_{t+1}(\tilde{\gamma})}{-\eta+(1-\alpha)\delta\rho^2\left((\eta-\zeta)\bar{\lambda}_{t+1}+\zeta\right)}. \tag{61}$$

We can analogously begin with the period t Euler equation as *predicted* before period t, which has $\tilde{\gamma}$ and $\bar{s}_t$ instead of γ and s_t on the left-hand side, and does not have the $(1-\alpha)$ term. This gives $\bar{x}_t = \bar{\lambda}_t\bar{s}_t + \mu_t(\tilde{\gamma})$, with

$$\bar{\lambda}_t = \frac{\zeta - \delta\rho^2\left((\eta-\zeta)\bar{\lambda}_{t+1}+\zeta\right)}{-\eta+\delta\rho^2\left((\eta-\zeta)\bar{\lambda}_{t+1}+\zeta\right)}. \tag{62}$$

and $\mu_t(\tilde{\gamma})$ given by equation (61) except that, as as implied by writing $\mu_t(\tilde{\gamma})$ instead of $\mu_t(\gamma)$, the third term in the numerator is $\tilde{\gamma}$ instead of γ.[25] Thus, λ_t is not correctly perceived in advance of period t.

λ_t depends only on $\{\eta,\zeta,\delta,\rho,\alpha\}$, and $\bar{\lambda}_t$ depends only on $\{\eta,\zeta,\delta,\rho\}$, as long as $\bar{\lambda}_{t+1}$ depends only on $\{\eta,\zeta,\delta,\rho\}$. Because consumers misperceive γ, μ_r is also misperceived for $r > t$. The function μ_t is linear in p_t. □

We now show that with concave utility, λ_t and $\bar{\lambda}_t$ are constant in t far from the time horizon.

Lemma 3. *Suppose the conditions for Lemma 2 hold and utility is concave. Then for any fixed t,*

$$\lambda = \lim_{T\to\infty}\lambda_t = \frac{\zeta - (1-\alpha)\delta\rho^2\left((\eta-\zeta)\bar{\lambda}+\zeta\right)}{-\eta+(1-\alpha)\delta\rho^2\left((\eta-\zeta)\bar{\lambda}+\zeta\right)}, \tag{63}$$

and

$$\bar{\lambda} = \lim_{T\to\infty}\bar{\lambda}_t = \frac{-\eta - \sqrt{\eta^2 - 4\frac{\delta\rho^2(\zeta-\eta)}{(1-\delta\rho^2)}\zeta}}{\frac{2\delta\rho^2(\zeta-\eta)}{(1-\delta\rho^2)}}. \tag{64}$$

Proof. To show that λ_t is constant in t far from the time horizon, it suffices to prove the convergence of $\bar{\lambda}_t$ to the steady state, since λ_t is a function of $\bar{\lambda}_{t+1}$ and other deterministic parameters. We define the function

[25] Equation (60) is much simpler than equation (25) of Gruber and Köszegi (2001), and our expression for λ_t does not depend on actual or perceived temptation γ or $\tilde{\gamma}$, while theirs depends on present focus β. This is because in their quasi-hyperbolic framework, $1-\beta$ multiplies λ_{t+1} parameters in the Euler equation and doesn't drop out.

$f(\tilde{\lambda})$ according to Equation (62) that describes the recursion $\tilde{\lambda}_t = f(\tilde{\lambda}_{t+1})$. We first find the values of $\tilde{\lambda}$ that could be fixed points. Assuming constant $\tilde{\lambda}$ and rearranging Equation (60) gives

$$-\eta\tilde{\lambda} + \delta\rho^2\left((\eta-\zeta)\tilde{\lambda}^2 + \zeta\tilde{\lambda}\right) = \zeta + \delta\rho^2\left((\zeta-\eta)-\zeta\right). \tag{65}$$

Collecting terms gives

$$\tilde{\lambda}^2\delta\rho^2(\eta-\zeta) + \tilde{\lambda}\eta\left(\delta\rho^2-1\right) + \zeta\left(\delta\rho^2-1\right) = 0 \tag{66}$$

$$\tilde{\lambda}^2\frac{\delta\rho^2(\zeta-\eta)}{(1-\delta\rho^2)} + \tilde{\lambda}\eta + \zeta = 0. \tag{67}$$

Using the quadratic formula gives

$$\tilde{\lambda} = \frac{-\eta \pm \sqrt{\eta^2 - 4\frac{\delta\rho^2(\zeta-\eta)}{(1-\delta\rho^2)}\zeta}}{\frac{2\delta\rho^2(\zeta-\eta)}{(1-\delta\rho^2)}}. \tag{68}$$

We now prove convergence. The function $f(\lambda)$ has the following properties. First, $f(\lambda)$ is always increasing as

$$f'(\tilde{\lambda}) = \frac{-\delta\rho^2(\eta-\zeta)\left(-\eta+\delta\rho^2\left((\eta-\zeta)\tilde{\lambda}+\zeta\right)\right) - \delta\rho^2(\eta-\zeta)\left(\zeta-\delta\rho^2\left((\eta-\zeta)\tilde{\lambda}+\zeta\right)\right)}{\left(-\eta+\delta\rho^2\left((\eta-\zeta)\tilde{\lambda}+\zeta\right)\right)^2} \tag{69}$$

$$= \frac{\delta\rho^2(\zeta-\eta)^2}{\left(-\eta+\delta\rho^2\left((\eta-\zeta)\tilde{\lambda}+\zeta\right)\right)^2} > 0. \tag{70}$$

Second, f is convex on $(-\infty, \bar{\tilde{\lambda}})$,where $\bar{\tilde{\lambda}} = \dfrac{-\eta+\delta\rho^2\zeta}{\delta\rho^2(\zeta-\eta)} > 0$. This comes from the sign of its second derivative

$$f''(\tilde{\lambda}) = \frac{2\delta^2\rho^4(-\eta+\zeta)^3}{\left(-\eta+\delta\rho^2\left((\eta-\zeta)\tilde{\lambda}+\zeta\right)\right)^3}, \tag{71}$$

which is determined by the sign of the denominator.

Third, for $\tilde{\lambda} > \bar{\tilde{\lambda}}$, $f(\tilde{\lambda})$ is always negative due to the denominator in equation (62), hence none of the solutions for a constant $\tilde{\lambda}_t$ are in this region.

Fourth, $f(0) > 0$ since $\delta\rho^2 < 1$ and

$$f(0) = \frac{\zeta(1-\delta\rho^2)}{-\eta+\delta\rho^2\zeta}. \tag{72}$$

Fifth, $f(\bar{\lambda})$ is continuous on $[0, \bar{\bar{\lambda}})$ and $\lim_{\bar{\lambda} \to \bar{\bar{\lambda}}} f(\bar{\lambda}) = \infty$ as the denominator in equation (60) goes to 0.

The properties highlighted above imply that both candidate solutions for a constant $\bar{\lambda}_t$ in equation (68) are positive. To see this, denote the two candidate solutions as $(\bar{\lambda}_1, \bar{\lambda}_2)$, with $\bar{\lambda}_1 < \bar{\lambda}_2$. Since $f(0) > 0$, we know that at least one solution for $\bar{\lambda}$ is positive given $-\eta > 0$. Furthermore, since $f(\bar{\lambda}) > 0$ on $(-\infty, \bar{\bar{\lambda}}]$, it cannot be true that an increasing, continuous, and convex function that diverges to infinity at $\bar{\bar{\lambda}}$ only crosses the identity function once in $[0, \bar{\bar{\lambda}})$. Hence, both solutions are in $[0, \bar{\bar{\lambda}}]$.

Given this result and the convex shape of this function, it must be true that $\bar{\lambda}_1$ is a stable constant solution for the recursion while $\bar{\lambda}_2$ is unstable. For any point in $[0, \bar{\lambda}_1]$ the recursion implies an increase in $\bar{\lambda}_t$ $(f(\bar{\lambda}) > \bar{\lambda})$, for any point in $[\bar{\lambda}_1, \bar{\lambda}_2]$ the recursion implies a decrease in λ_t $(f(\bar{\lambda}) < \bar{\lambda})$, and for any point in $[\bar{\lambda}_2, \bar{\bar{\lambda}}]$ the recursion implies an increase in $\bar{\lambda}_t$ $(f(\bar{\lambda}) > \bar{\lambda})$. Overall, this means that for any starting value of $\bar{\lambda}_t \in [0, \bar{\lambda}_2)$ the recursion converges to $\bar{\lambda}_1$.

To complete the proof, we begin with $\bar{\lambda}_T$ and then prove that far away from the time horizon, $\bar{\lambda}_t$ is constant. To do this, we need to show that this initial value, given by $\bar{\lambda}_T = \dfrac{\zeta}{-\eta}$, is less than $\bar{\lambda}_2$. To show this, notice that the two solutions $(\bar{\lambda}_1, \bar{\lambda}_2)$ are symmetrically placed around $\bar{\lambda}_s = \dfrac{-\eta(1-\delta\rho^2)}{2\delta\rho^2(\zeta-\eta)}$. Given this value, by the parametric assumption that guarantees the existence of the two constant solutions for the recursion, we know that

$$\eta^2 - 4\frac{\delta\rho^2(\zeta-\eta)}{(1-\delta\rho^2)}\zeta > 0, \tag{73}$$

and since

$$\eta^2 > 2\frac{\delta\rho^2(\zeta-\eta)\zeta}{(1-\delta\rho^2)} \iff \frac{\zeta}{-\eta} < \frac{-\eta(1-\delta\rho^2)}{2\delta\rho^2(\zeta-\eta)}, \tag{74}$$

we have that $\bar{\lambda}_T < \bar{\lambda}_s$. Then $\bar{\lambda}_T < \bar{\lambda}_s < \lambda_2$, and hence the backward recursion starting from $\bar{\lambda}_T$ converges far from the time horizon to a stationary value $\bar{\lambda}^* = \bar{\lambda}_1$. Moreover, $f(\bar{\lambda}_T)$ can be written as $\frac{\zeta - X}{-\eta + X}$, and we know that $\frac{\zeta - X}{-\eta + X} > \frac{\zeta}{-\eta}$ whenever $X < 0$. Then, given that

$$X = (1-\alpha)\delta\rho^2\Big((\eta-\zeta)\bar{\lambda}_T + \zeta\Big) < 0 \iff (\eta-\zeta)\frac{\zeta}{-\eta} + \zeta < 0 \iff \frac{\zeta^2}{\eta} < 0 \iff \eta < 0,$$

we have $X < 0$. Thus we can conclude that $f(\bar{\lambda}_T) > \bar{\lambda}_T$ and therefore, $\bar{\lambda}_T < \bar{\lambda}_1$. Thus, we have proved that the backward recursion converges to an stationary value of $\bar{\lambda}^* = \bar{\lambda}_1$, and it does so as an increasing sequence.

Finally, we demonstrate that λ_t also converges to a steady-state in a decreasing manner. We note that

$$\lambda = g(\bar{\lambda}) = \frac{\zeta - (1-\alpha)\delta\rho^2\Big((\eta-\zeta)\bar{\lambda} + \zeta\Big)}{-\eta + (1-\alpha)\delta\rho^2\Big((\eta-\zeta)\bar{\lambda} + \zeta\Big)} \tag{75}$$

Which we can rewrite as

$$\lambda = g(\tilde{\lambda}) = \frac{\zeta + (1-\alpha)\delta\rho^2\left((\zeta-\eta)\tilde{\lambda}+\zeta\right)}{-\eta-(1-\alpha)\delta\rho^2\left((\zeta-\eta)\tilde{\lambda}+\zeta\right)} \tag{76}$$

Note that $(1-\alpha)\delta\rho^2(\zeta-\eta)$ is positive, so the numerator decreases when $\tilde{\lambda}$ decreases, whereas the denominator increases, since $-(1-\alpha)\delta\rho^2(\zeta-\eta)\tilde{\lambda}$ becomes less negative. Hence, $g(\tilde{\lambda}) = \lambda$ also decreases when $\tilde{\lambda}$ decreases. $\square$

We now show that utility is concave in x_t as long as there is not too much habit formation in a specific sense.

Lemma 4. *Suppose the conditions for Lemma 2 hold and U_t is given by equation (37). Then for any t, $\frac{dU_t}{dx_t}$ is continuous in x_t. Furthermore, if $\tilde{\lambda}^b$ is an upper bound on $\tilde{\lambda}_t$ and $\frac{(1-\alpha)\tilde{\lambda}^b}{(1+\tilde{\lambda}^b)-\delta\rho^2(1+\tilde{\lambda}^b)^2} < \frac{-\eta}{\zeta}$, then $\frac{\partial^2 U_t}{\partial x_t^2} < 0$ for all $t \geq 0$.*

Proof. The period t decisionmaker maximizes equation (37). The derivative of equation (37) can be written as

$$\frac{dU_t\left(x_t;s_t\right)}{dx_t} = \frac{\partial u_t}{\partial x_t} + \gamma + (1-\alpha)\sum_{r=t+1}^{T}\delta^{r-t}\frac{\partial \tilde{s}_r}{\partial x_t}\left[\underbrace{\frac{\partial \tilde{u}_r}{\partial \tilde{x}_r}\frac{\partial \tilde{x}_r}{\partial \tilde{s}_r} + \frac{\partial \tilde{u}_r}{\partial \tilde{s}_r}}_{\text{effect of } \tilde{s}_r \text{ on period } r \text{ utility}} + \underbrace{\delta\rho\frac{\partial V_{r+1}}{\partial \tilde{s}_{r+1}}\frac{\partial \tilde{x}_r}{\partial \tilde{s}_r}}_{\text{partial effect on future utility}}\right]. \tag{77}$$

The summation term in equation (77) is the effect on future utility from the change in habit stock brought into future periods. $\frac{\partial \tilde{s}_r}{\partial x_t} = \rho^{r-t}$ is the predicted direct effect of consumption $\tilde{x}_t$ on stock in period r. The first two terms inside brackets are the effect of that change on period r utility. The final term inside brackets accounts for the fact that the resulting change in $\tilde{x}_r$ will affect utility in later periods.

The period t decisionmaker predicts that her period $r > t$ self will maximize equation (36). The predicted period r first-order condition is

$$\left.\frac{d\tilde{U}_r\left(x_r;\tilde{s}_r\right)}{d\tilde{x}_r}\right|_{\tilde{x}_r} = 0 = \frac{\partial \tilde{u}_r}{\partial \tilde{x}_r} + \tilde{\gamma} + \delta\rho\frac{\partial V_{r+1}}{\partial \tilde{s}_{r+1}}. \tag{78}$$

Multiplying this FOC by $\tilde{\lambda}_r := \frac{\partial \tilde{x}_r}{\partial \tilde{s}_r}$ and subtracting it from the term inside brackets in equation (77) gives

$$\frac{dU_t}{dx_t} = \frac{\partial u_t}{\partial x_t} + \gamma + (1-\alpha)\sum_{r=t+1}^{T} \delta^{r-t}\rho^{r-t} \left[\begin{array}{c} \frac{\partial \bar{u}_r}{\partial \bar{x}_r}\bar{\lambda}_r + \frac{\partial \bar{u}_r}{\partial \bar{s}_r} + \delta\rho \frac{\partial V_{r+1}}{\partial \bar{s}_{r+1}}\bar{\lambda}_r \\ -\left[\frac{\partial \bar{u}_r}{\partial \bar{x}_r}\bar{\lambda}_r + \tilde{\gamma}\bar{\lambda}_r + \delta\rho\frac{\partial V_{r+1}}{\partial \bar{s}_{r+1}}\bar{\lambda}_r\right] \end{array} \right] \tag{79}$$

$$= \frac{\partial u_t}{\partial x_t} + \gamma + (1-\alpha)\sum_{r=t+1}^{T} (\delta\rho)^{r-t}\left[\frac{\partial \bar{u}_r}{\partial \bar{s}_r} - \tilde{\gamma}\bar{\lambda}_r\right] \tag{80}$$

With the quadratic functional form, this becomes

$$\frac{dU_t}{dx_t} = \eta x_t + \zeta s_t + \xi_t - p_t + \gamma + (1-\alpha)\sum_{r=t+1}^{T} (\delta\rho)^{r-t}\left[\zeta\bar{x}_r + \phi - \tilde{\gamma}\bar{\lambda}\right]. \tag{81}$$

In this equation, two terms (x_t and $\bar{x}_r$) depend on x_t. x_t is by definition continuous in x_t , and $\bar{x}_r$ is continuous in past consumption x_t due to the evolution of habit stock and Lemma 2. Thus, $\frac{dU_t}{dx_t}$ is continuous in x.

We now turn to concavity. The derivative of equation (81) is

$$\frac{d^2U_t}{dx_t^2} = \eta + (1-\alpha)\sum_{r=t+1}^{\infty} (\delta\rho)^{r-t}\zeta\frac{d\bar{x}_r}{dx_t}. \tag{82}$$

$$= \eta + (1-\alpha)\sum_{r=t+1}^{\infty} (\delta\rho)^{r-t}\zeta\bar{\lambda}_r\left[\rho^{r-t}\prod_{j=t+1}^{r-1}\left(1+\bar{\lambda}_j\right)\right] \tag{83}$$

Intuitively, $\frac{d^2U_t}{dx_t^2} < 0$ requires that the diminishing marginal utility in period t outweighs the incentive to increase current consumption for the purpose of increasing future utility through ζ. This will tend to be true when projection bias α is large and/or habit formation ρ is small. A small ρ has a direct effect by causing the habit stock from dx_t to decay faster. It also has an indirect effect by reducing $\frac{d\bar{x}_r}{dx_t}$, the perceived effect of current consumption on future consumption.

If we know an upper bound $\bar{\lambda}^b$ such that $\bar{\lambda}^b > \bar{\lambda}_t$ for all t, we can write a simpler necessary condition

for concavity: $\frac{d^2U_t}{d\bar{x}^2} < 0$ for all $t \geq 0$ if

$$(1-\alpha)\sum_{r=t+1}^{\infty}(\delta\rho)^{r-t}\bar{\lambda}_r\left[\rho^{r-t}\prod_{j=t+1}^{r-1}\left(1+\bar{\lambda}_j\right)\right] < \frac{-\eta}{\zeta} \quad (84)$$

$$(1-\alpha)\sum_{r=t+1}^{\infty}(\delta\rho)^{r-t}\bar{\lambda}^b\left[\rho^{r-t}\left(1+\bar{\lambda}^b\right)^{r-t-1}\right] < \frac{-\eta}{\zeta} \quad (85)$$

$$(1-\alpha)\frac{\bar{\lambda}^b}{1+\bar{\lambda}^b}\cdot\sum_{r=1}^{\infty}\left(\delta\rho^2\left(1+\bar{\lambda}^b\right)\right)^{r-1} < \frac{-\eta}{\zeta} \quad (86)$$

$$(1-\alpha)\frac{\bar{\lambda}^b}{1+\bar{\lambda}^b}\cdot\left[\frac{1}{1-\left(\delta\rho^2\left(1+\bar{\lambda}^b\right)\right)}\right] < \frac{-\eta}{\zeta} \quad (87)$$

$$\frac{(1-\alpha)\bar{\lambda}^b}{\left(1+\bar{\lambda}^b\right)-\delta\rho^2\left(1+\bar{\lambda}^b\right)^2} < \frac{-\eta}{\zeta}. \quad (88)$$

□

From the proof of Lemma 3, we know that $\bar{\lambda}_t$ decreases as $t \to T$.

Finally, we show the conditions under which μ_t converges to a constant far from the time horizon.

Lemma 5. *Suppose the conditions for Lemma 2 hold, and* $-\eta > (1-\alpha)\delta\rho\left[(\zeta-\eta)\left(1+\rho\bar{\lambda}_{t+1}\right)-\rho\zeta\right]$. *Then* $\lim_{(T-t)\to\infty}\mu_t = \mu$.

Proof. Since $\mu_t(\gamma)$ is a function of only constants, $\bar{\lambda}_{t+1}$ (which converges per Lemma 3), and $\mu_{t+1}(\bar{\gamma})$, it is sufficient to show that the sequence $\mu_t(\bar{\gamma})$ converges. The coefficient on $\mu_{t+1}(\bar{\gamma})$ in equation (61) is

$$\frac{(1-\alpha)\delta\rho(\zeta-\eta)}{-\eta+(1-\alpha)\delta\rho^2\left((\eta-\zeta)\bar{\lambda}_{t+1}+\zeta\right)}. \quad (89)$$

The sequence $\mu_{t+1}(\bar{\gamma})$ will converge if and only if

$$\frac{(1-\alpha)\delta\rho(\zeta-\eta)}{-\eta+(1-\alpha)\delta\rho^2\left((\eta-\zeta)\bar{\lambda}_{t+1}+\zeta\right)} < 1. \quad (90)$$

The denominator is positive at our parameter values, so this inequality requires

$$-\eta > (1-\alpha)\delta\rho\left[(\zeta-\eta)\left(1+\rho\bar{\lambda}_{t+1}\right)-\rho\zeta\right]. \quad (91)$$

In words, this requires that perceived habit formation $(1-\alpha)\rho$ is small relative to the demand slope parameter η. □

Proposition 2 combines Lemmas 2, 3, 4, and 5.

F.3 Proof of Lemma 1: Steady-State Convergence

Proof. Capital stock evolves according to $s_t = \rho(s_{t-1} + x_{t-1})$. Substituting in the stable equilibrium strategy $x_t^* = \lambda s_t + \mu$ gives

$$s_t = \rho(s_{t-1} + \lambda s_{t-1} + \mu) \tag{92}$$
$$= \rho\mu + \rho(1+\lambda)s_{t-1} \tag{93}$$
$$= \rho\mu + \rho(1+\lambda)(\rho\mu + \rho(1+\lambda)s_{t-2}) \tag{94}$$
$$= \rho\mu + \rho^2(1+\lambda)\mu + \rho^2(1+\lambda)^2 s_{t-2} \tag{95}$$
$$= \rho\mu + \rho^2(1+\lambda)\mu + \rho^3(1+\lambda)^2\mu + \rho^3(1+\lambda)^3 s_{t-3}. \tag{96}$$

Thus

$$s_t = \frac{\mu}{1+\lambda}\left(\iota + \iota^2 + \dots + \iota^k\right) + \iota^k s_{t-k}, \tag{97}$$

where $\iota = (1+\lambda)\rho$. Thus, provided that $\iota < 1$, in the limit as $k \to \infty$ we have

$$s_t = \frac{\mu}{1+\lambda} \cdot \frac{\iota}{1-\iota} \tag{98}$$
$$= \frac{\mu\rho}{1-(1+\lambda)\rho}. \tag{99}$$

We can then check that this is indeed a steady state:

$$s_t = \rho\left(\frac{\mu\rho}{1-(1+\lambda)\rho} + \mu + \lambda\left(\frac{\mu\rho}{1-(1+\lambda)\rho}\right)\right) \tag{100}$$
$$= \rho\left(\frac{\mu\rho + \mu(1-(1+\lambda)\rho) + \lambda\mu\rho}{1-(1+\lambda)\rho}\right) \tag{101}$$
$$= \rho\left(\frac{\mu\rho + \mu - \mu\rho - \mu\lambda\rho + \lambda\mu\rho}{1-(1+\lambda)\rho}\right) \tag{102}$$
$$= \frac{\mu\rho}{1-(1+\lambda)\rho} \tag{103}$$

□

F.4 Proof of Proposition 3: Steady-State Consumption

Proof. We assume steady state implies constant consumption and habit stock, but not necessarily constant *predicted* consumption and habit stock. In steady state, $p_t = p$, $\xi_t = \xi$, $s_t = s_{ss}$, and $x_t = x_{ss}$. By equation (1)

governing the evolution of habit stock, $s_{ss} = \rho(s_{ss} + x_{ss})$, and re-arranging this equation gives $s_{ss} = \frac{\rho}{1-\rho} x_{ss}$. Earlier, we defined steady-state misprediction as $m_{ss} := \tilde{x}_{t+1} - x_{ss}$.

We substitute $p_t = p$, $\xi_t = \xi$, $s_t = s_{ss}$, and $x_t = x_{ss}$ into the Euler equation (equation (18)), giving

$$\eta x_{ss} + \zeta s_{ss} + \xi - p + \gamma = (1-\alpha)\delta\rho \left[\eta \tilde{x}_{t+1} + \zeta\rho\left(x_{ss} + s_{ss}\right) + \xi - p + \left(1+\tilde{\lambda}\right)\tilde{\gamma} - \zeta\tilde{x}_{t+1} - \phi\right]. \tag{104}$$

Substituting in $s_{ss} = \frac{\rho}{1-\rho} x_{ss}$ and also writing predicted consumption as a deviation from the actual value gives

$$\eta x_{ss} + \xi - p + \frac{\rho\zeta}{1-\rho} x_{ss} + \gamma = (1-\alpha)\delta\rho \left[(\eta - \zeta)\left((\tilde{x}_{t+1} - x_{ss}) + x_{ss}\right) + \zeta\rho\left(\frac{1}{1-\rho} x_{ss}\right) + \xi - p + \left(1+\tilde{\lambda}\right)\tilde{\gamma} - \phi\right]. \tag{105}$$

Substituting $m_{ss} := \tilde{x}_{t+1} - x_{ss}$ and collecting terms gives

$$\begin{aligned} x_{ss}\left[\eta + \frac{\rho\zeta}{1-\rho} - (1-\alpha)\delta\rho\left((\eta-\zeta) + \frac{\zeta\rho}{1-\rho}\right)\right] &= p - \xi - \gamma + (1-\alpha)\delta\rho\Big[(\eta-\zeta)m_{ss} \\ &\quad + \xi - p + \left(1+\tilde{\lambda}\right)\tilde{\gamma} - \phi\Big] \qquad (106) \\ x_{ss}\left[\eta - (1-\alpha)\delta\rho(\eta-\zeta) + \zeta\frac{\rho - (1-\alpha)\delta\rho^2}{1-\rho}\right] &= \left(1 - (1-\alpha)\delta\rho\right)(p - \xi) + (1-\alpha)\delta\rho\Big[(\eta-\zeta)m_{ss} \\ &\quad + \left(1+\tilde{\lambda}\right)\tilde{\gamma} - \phi\Big] - \gamma. \qquad (107) \end{aligned}$$

Multiplying both sides by (-1), setting $\kappa := (1-\alpha)\delta\rho(\phi - \xi) + \xi$, and dividing through gives equation (21).

□

G Derivations of Estimating Equations in Appendix E.3

We define $y^g := \mathbb{E}_{i \in g} y_i$ as the expectation over individuals in group g of parameter y. Due to random assignment, $\xi_{st}^g = \xi_{st}^{g'}$ and $s_2^g = s_2^{g'}$ for all $\{g, g'\}$, and $\mu_t^B = \mu_t^{BC}$ for $t \in \{2,4,5\}$. The estimating equations for the restricted model in Section 6.2 are the below equations with the additional assumptions that $\tau_2^B = 0$ and $\alpha = 1$.

G.1 Habit Formation

Derivation of equation (22). From equation (19) and the evolution of habit stock, we have

$$x_4^* = \lambda s_4 + \mu_4 \tag{108}$$
$$= \lambda\rho\,(s_3 + x_3^*) + \mu_4 \tag{109}$$
$$= \lambda\rho\,(\rho\,(s_2 + x_2^*) + x_3^*) + \mu_4. \tag{110}$$

Thus, group average consumption is $x_4^g = \lambda\left(\rho^2\left(s_2^g + x_2^g\right) + \rho x_3^g\right) + \mu_4^g$, and the period 4 bonus effect is

$$\tau_4^B = \lambda\left(\rho^2\tau_2^B + \rho\,\tau_3^B\right). \tag{111}$$

Re-arranging gives equation (22).

Derivation of equation (23). Similarly, we have

$$x_5^* = \lambda s_5 + \mu_5 \tag{112}$$
$$= \lambda\rho\,(s_4 + x_4^*) + \mu_5 \tag{113}$$
$$= \lambda\rho\,(\rho\,(s_3 + x_3^*) + x_4^*) + \mu_5 \tag{114}$$
$$= \lambda\rho\,(\rho\,(\rho\,(s_2 + x_2^*) + x_3^*) + x_4^*) + \mu_5. \tag{115}$$

Thus, group average consumption is $x_5^g = \lambda\left(\rho^3\left(s_2^g + x_2^g\right) + \rho^2 x_3^g + \rho x_4^g\right) + \mu_5^g$, and the period 5 bonus effect is

$$\tau_5^B = \lambda\left(\rho^3\tau_2^B + \rho^2\tau_3^B + \rho\,\tau_4^B\right). \tag{116}$$

Multiplying equation (111) by ρ and subtracting from equation (116) gives $\tau_5^B - \tau_4^B\rho = \lambda\rho\,\tau_4^B$, and re-arranging gives equation (23).

System of equations for λ and ρ. Re-arranging equation (23) gives

$$\lambda = \frac{\tau_5^B}{\tau_4^B\rho} - 1. \tag{117}$$

Substituting this into equation (22) gives:

$$\frac{\tau_5^B - \tau_4^B\rho}{\tau_4^B\rho} = \frac{\tau_4^B}{\rho\,\tau_3^B + \rho^2\tau_2^B} \tag{118}$$
$$\left(\tau_4^B\right)^2 = \left(\tau_5^B - \tau_4^B\rho\right)\left(\tau_3^B + \rho\,\tau_2^B\right) \tag{119}$$
$$0 = \left[\tau_2^B\tau_4^B\right]\rho^2 + \left[\tau_3^B\tau_4^B - \tau_2^B\tau_5^B\right]\rho + \left[\left(\tau_4^B\right)^2 - \tau_3^B\tau_5^B\right]. \tag{120}$$

The quadratic formula gives

$$\rho = \frac{-\left[\tau_3^B\tau_4^B - \tau_2^B\tau_5^B \pm \sqrt{\left[\tau_3^B\tau_4^B - \tau_2^B\tau_5^B\right]^2 - 4\left[\tau_2^B\tau_4^B\right]\left[\left(\tau_4^B\right)^2 - \tau_3^B\tau_5^B\right]}\right]}{2\left[\tau_2^B\tau_4^B\right]} \tag{121}$$

In all bootstrap draws in our data, only one of the two solutions satisfies the requirement that $\rho \geq 0$.

Special case with $\tau_2^B = 0$. If there is no anticipatory demand response ($\tau_2^B = 0$), we have $\tau_4^B = \lambda\rho\tau_3^B$ and $\tau_5^B = \lambda\rho^2\tau_3^B + \lambda\rho\tau_4^B$. Dividing the two equations gives

$$\frac{\tau_5^B}{\tau_4^B} = \rho + \frac{\tau_4^B}{\tau_3^B}$$
$$\rho = \frac{\tau_5^B}{\tau_4^B} - \frac{\tau_4^B}{\tau_3^B}. \tag{122}$$

We then solve for λ by inserting this ρ into equation (22) with $\tau_2^B = 0$.

G.2 Perceived Habit Formation, Price Response, and Habit Stock Effect on Marginal Utility

The expectation over i of the Euler equations for group g is

$$\eta x_t^g + \zeta s_t^g + \xi_t^g - p_t + \gamma = (1-\alpha)\delta\rho\left[\eta\tilde{x}_{t+1}^g + \zeta\tilde{s}_{t+1}^g + \xi_{t+1}^g - p_{t+1} + \tilde{\gamma} + \tilde{\gamma}\tilde{\lambda}_{t+1} - \left(\zeta\tilde{x}_{t+1}^g + \phi\right)\right] \tag{123}$$

Derivation of equation (24). Differencing the Euler equations for periods 2 versus 3 for the Bonus and Bonus Control groups gives

$$\eta\tau_2^B = (1-\alpha)\delta\rho\left[-p^B + (\eta-\zeta)\left(\tilde{x}_3^B - \tilde{x}_3^{BC}\right) + \zeta\left(\tilde{s}_3^B - \tilde{s}_3^{BC}\right)\right]. \tag{124}$$

Substituting $\tilde{x}_3^B - \tilde{x}_3^{BC} = \tilde{\tau}_3^B$ and $\tilde{s}_3^B - \tilde{s}_3^{BC} = \rho\tau_2^B$ gives

$$\eta\tau_2^B = (1-\alpha)\delta\rho\left[-p^B + (\eta-\zeta)\tilde{\tau}_3^B + \zeta\rho\tau_2^B\right]. \tag{125}$$

Rearranging gives equation (24).

If $\tilde{\gamma} \neq \gamma$, then people update their predictions of $\tilde{x}_3$ as they set x_2^*, and thus the predictions of $\tilde{x}_3$ from survey 2 are inconsistent with x_2^*. However, there is only limited misprediction in our data, so this is not very consequential.

Derivation of equation (25). Differencing the Euler equations for periods 3 versus 4 for the Bonus and Bonus Control groups gives

$$\left(-p^{B}-0\right)+\eta\,\tau_3^{B}+\zeta\left(s_3^{B}-s_3^{BC}\right)=(1-\alpha)\delta\rho\left[(\eta-\zeta)\left(\bar{x}_4^{B}-\bar{x}_4^{BC}\right)+\zeta\left(\bar{s}_4^{B}-\bar{s}_4^{BC}\right)\right]. \tag{126}$$

Habit stock evolution implies $s_3^B-s_3^{BC}=\rho\left(s_2^B-s_2^{BC}+x_2^B-x_2^{BC}\right)=\rho\,\tau_2^B$ and $\bar{s}_4^B-\bar{s}_4^{BC}=\rho\left(s_3^B-s_3^{BC}+x_3^B-x_3^{BC}\right)=\rho\left(\rho\tau_2^B+\tau_3^B\right)$. Linear policy functions imply $\bar{x}_4=\bar{\lambda}\bar{s}_4+\bar{\mu}_4$, so $\bar{x}_4^B-\bar{x}_4^{BC}=\bar{\lambda}\left(\bar{s}_4^B-\bar{s}_4^{BC}\right)$. Substituting these equations gives

$$\left(-p^{B}-0\right)+\eta\,\tau_3^{B}+\zeta\rho\,\tau_2^{B}=(1-\alpha)\delta\rho\left[\left((\eta-\zeta)\bar{\lambda}+\zeta\right)\rho\left(\rho\tau_2^{B}+\tau_3^{B}\right)\right]. \tag{127}$$

Rearranging gives

$$\eta\left(\tau_3^{B}-(1-\alpha)\delta\rho^2\bar{\lambda}\left(\rho\tau_2^{B}+\tau_3^{B}\right)\right)=p^{B}-\zeta\rho\,\tau_2^{B}+(1-\alpha)\delta\rho^2\zeta\left(1-\bar{\lambda}\right)\left(\rho\tau_2^{B}+\tau_3^{B}\right). \tag{128}$$

Solving for η gives equation (25).

Derivation of equation (26). Differencing the Euler equations for periods 4 versus 5 for the Bonus and Bonus Control groups gives

$$\eta\left(x_4^{B}-x_4^{BC}\right)+\zeta\left(s_4^{B}-s_4^{BC}\right)=(1-\alpha)\delta\rho\left[(\eta-\zeta)\left(\bar{x}_5^{B}-\bar{x}_5^{BC}\right)+\zeta\left(\bar{s}_5^{B}-\bar{s}_5^{BC}\right)\right] \tag{129}$$

Habit stock evolution implies $s_4^B-s_4^{BC}=\rho\left(s_3^B-s_3^{BC}+x_3^B-x_3^{BC}\right)=\rho^2\tau_2^B+\rho\,\tau_3^B$ and $\bar{s}_5^B-\bar{s}_5^{BC}=\rho\left(s_4^B-s_4^{BC}+x_4^B-x_4^{BC}\right)=\rho\left(\rho^2\tau_2^B+\rho\tau_3^B+\tau_4^B\right)$. Linear policy functions imply $\bar{x}_5=\bar{\lambda}\bar{s}_5+\bar{\mu}_5$, so $\bar{x}_5^B-\bar{x}_5^{BC}=\bar{\lambda}\left(\bar{s}_5^B-\bar{s}_5^{BC}\right)$. Substituting these equations gives

$$\eta\,\tau_4^{B}+\zeta\left(\rho^2\tau_2^{B}+\rho\,\tau_3^{B}\right)=(1-\alpha)\delta\rho\left[\left((\eta-\zeta)\bar{\lambda}+\zeta\right)\rho\left(\rho^2\tau_2^{B}+\rho\,\tau_3^{B}+\tau_4^{B}\right)\right] \tag{130}$$

$$=(1-\alpha)\delta\rho^2\left[\left(\eta\lambda+\zeta\left(1-\bar{\lambda}\right)\right)\left(\rho^2\tau_2^{B}+\rho\,\tau_3^{B}+\tau_4^{B}\right)\right]. \tag{131}$$

Collecting ζ terms gives

$$\zeta\left(\rho\tau_3^{B}+\rho^2\tau_2^{B}\right)-(1-\alpha)\delta\rho^2\left[\zeta\left(1-\bar{\lambda}\right)\left(\rho^2\tau_2^{B}+\rho\,\tau_3^{B}+\tau_4^{B}\right)\right]=-\eta\,\tau_4^{B}+(1-\alpha)\delta\rho^2\eta\bar{\lambda}\left(\rho^2\tau_2^{B}+\rho\,\tau_3^{B}+\tau_4^{B}\right). \tag{132}$$

Solving for ζ gives equation (26).

System of equations for $(1-\alpha)$, η, and ζ.

First, we solve explicitly for $(1-\alpha)$ before substituting it back in Equations (25) and (26) to solve for η and ζ.

We define

$$y := \frac{-\tau_4^B + (1-\alpha)\delta\rho^2\lambda\left[\rho^2\tau_2^B + \rho\tau_3^B + \tau_4^B\right]}{\rho\tau_3^B + \rho^2\tau_2^B - (1-\alpha)\delta\rho^2(1-\lambda)\left[\rho^2\tau_2^B + \rho\tau_3^B + \tau_4^B\right]}. \tag{133}$$

Observe that

$$\zeta = \eta \cdot y. \tag{134}$$

We can use this observation to rearrange Equation (25):

$$\eta = \frac{p^B - \zeta\rho\tau_2^B + (1-\alpha)\delta\rho^2\zeta(1-\lambda)\left(\rho\tau_2^B + \tau_3^B\right)}{\tau_3^B - (1-\alpha)\delta\rho^2\lambda\left(\rho\tau_2^B + \tau_3^B\right)} \tag{135}$$

$$\eta\left[\tau_3^B - (1-\alpha)\delta\rho^2\lambda\left(\rho\tau_2^B + \tau_3^B\right)\right] = p^B - \zeta\left(\rho\tau_2^B - (1-\alpha)\delta\rho^2(1-\lambda)\left(\rho\tau_2^B + \tau_3^B\right)\right) \tag{136}$$

$$= p^B - \eta\cdot y\left(\rho\tau_2^B - (1-\alpha)\delta\rho^2(1-\lambda)\left(\rho\tau_2^B + \tau_3^B\right)\right) \tag{137}$$

$$p^B = \eta\left[\tau_3^B - (1-\alpha)\delta\rho^2\lambda\left(\rho\tau_2^B + \tau_3^B\right) + y\left(\rho\tau_2^B - (1-\alpha)\delta\rho^2(1-\lambda)\left(\rho\tau_2^B + \tau_3^B\right)\right)\right]. \tag{138}$$

Then, define

$$x := \tau_3^B - (1-\alpha)\delta\rho^2\lambda\left(\rho\tau_2^B + \tau_3^B\right) + y\left(\rho\tau_2^B - (1-\alpha)\delta\rho^2(1-\lambda)\left(\rho\tau_2^B + \tau_3^B\right)\right) \tag{139}$$

where we observe that

$$\eta = \frac{p^B}{x}, \tag{140}$$

and

$$\zeta = \frac{p^B y}{x}. \tag{141}$$

Finally, we get that

$$(1-\alpha) = \frac{\eta\tau_2^B}{\delta\rho\left[-p^B + (\eta-\zeta)\tau_3^B + \zeta\rho\tau_2^B\right]} \tag{142}$$

$$= \frac{\frac{p^B}{x}\tau_2^B}{\delta\rho\left[-p^B + \left(\frac{p^B}{x} - \frac{p^B y}{x}\right)\tau_3^B + \frac{p^B y}{x}\rho\tau_2^B\right]}. \tag{143}$$

Since all scalars are known in the last equation, we can now solve for α. Then, we can estimate η and ζ by substituting α in Equations (25) and (26) respectively.

G.3 Naivete about Temptation

Derivation of equation (27). The Euler equation *predicted* for period t on the survey at the beginning of period t is

$$\eta x_t^*(s_t, \tilde{\gamma}, p_t) + \zeta s_t + \xi_t - p_t + \tilde{\gamma} = (1-\alpha)\delta\rho\left[\eta \tilde{x}_{t+1} + \zeta s_{t+1} + \xi_{t+1} - p_{t+1} + \tilde{\gamma} + \tilde{\gamma}\tilde{\lambda} - (\zeta \tilde{x}_{t+1} + \phi)\right]. \tag{144}$$

This equation uses the assumption that consumers are aware of period t projection bias when predicting period t consumption on survey t, so the only reason why the period t survey-taker mispredicts the period t objective function is naivete about period t temptation.

Habit stock evolution implies $\tilde{s}_{t+1} = \rho(s_t + \tilde{x}_t)$. Linear policy functions imply $\tilde{x}_{t+1} = \tilde{\lambda}\tilde{s}_{t+1} + \tilde{\mu}_{t+1}$. Substituting these equations into the predicted Euler equation gives

$$\eta x_t^*(s_t, \tilde{\gamma}, p_t) + \zeta s_t + \xi_t - p_t + \tilde{\gamma} = (1-\alpha)\delta\rho\left[(\eta-\zeta)\left(\tilde{\lambda}\tilde{s}_{t+1} + \tilde{\mu}_{t+1}\right) + \zeta\tilde{s}_{t+1} + \xi_{t+1} - p_{t+1} + \tilde{\gamma} + \tilde{\gamma}\tilde{\lambda} - \phi\right]. \tag{145}$$

$$= (1-\alpha)\delta\rho\left[\left((\eta-\zeta)\tilde{\lambda} + \zeta\right)\tilde{s}_{t+1} + (\eta-\zeta)\tilde{\mu}_{t+1} + \xi_{t+1} - p_{t+1} + \tilde{\gamma} + \tilde{\gamma}\tilde{\lambda} - \phi\right] \tag{146}$$

$$= (1-\alpha)\delta\rho\left[\left((\eta-\zeta)\tilde{\lambda} + \zeta\right)\rho(s_t + \tilde{x}_t) + (\eta-\zeta)\tilde{\mu}_{t+1} + \xi_{t+1} - p_{t+1} + \tilde{\gamma} + \tilde{\gamma}\tilde{\lambda} - \phi\right] \tag{147}$$

Analogously, the actual Euler equation for period t can be written as

$$\eta x_t^*(s_t, \gamma, p_t) + \zeta s_t + \xi_t - p_t + \gamma = (1-\alpha)\delta\rho\left[\left((\eta-\zeta)\tilde{\lambda} + \zeta\right)\rho(s_t + x_t^*) + (\eta-\zeta)\tilde{\mu}_{t+1} + \xi_{t+1} - p_{t+1} + \tilde{\gamma} + \tilde{\gamma}\tilde{\lambda} - \phi\right]. \tag{148}$$

Differencing the actual and predicted Euler equations for period t versus period $t+1$ for the Control group gives

$$\eta\left(x_t^C - \tilde{x}_t^C\right) + \gamma - \tilde{\gamma} = (1-\alpha)\delta\rho\left[\left((\eta-\zeta)\tilde{\lambda} + \zeta\right)\rho\left(x_t^C - \tilde{x}_t^C\right)\right] \tag{149}$$

Solving for $\gamma - \tilde{\gamma}$ and substituting $m^C = x_t^C - \tilde{x}_t^C$ gives equation (27).

G.4 Temptation

Limit effect: derivation of equation (28). Consider a "zero temptation" intervention that fully eliminates both perceived and actual temptation starting in period 2, generating treatment effects τ_t^0. Differencing the average Euler equations for periods 2 versus 3 for the zero temptation group versus its control group gives

$$\eta\left(x_2^0 - x_2^{0C}\right) - \gamma = (1-\alpha)\delta\rho\left[(\eta-\zeta)\left(\tilde{x}_3^0 - \tilde{x}_3^{0C}\right) + \zeta\left(\tilde{s}_3^0 - \tilde{s}_3^{0C}\right) - \tilde{\gamma} - \tilde{\gamma}\tilde{\lambda}\right] \tag{150}$$

$$\eta\tau_2^0 - \gamma = (1-\alpha)\delta\rho\left[(\eta-\zeta)\tilde{\tau}_3^0 + \zeta\rho\tau_2^0 - \tilde{\gamma} - \tilde{\gamma}\tilde{\lambda}\right] \tag{151}$$

Solving for γ and substituting $\tau^0 = \tau^L/\omega$ gives equation (28).

To solve for γ as a function of data and known parameters, we solve equation (27) for $\tilde{\gamma}$, substitute into equation (28), and rearrange, giving

$$\gamma = \frac{\eta\tau_2^L/\omega - (1-\alpha)\delta\rho\left(\left[(\eta-\zeta)\tilde{\tau}_3^L/\omega + \zeta\rho\tau_2^L/\omega\right] + (1+\tilde{\lambda})m_2^C\cdot\left[-\eta + (1-\alpha)\delta\rho^2\left((\eta-\zeta)\tilde{\lambda} + \zeta\right)\right]\right)}{1-(1-\alpha)\delta\rho(1+\tilde{\lambda})}. \tag{152}$$

Bonus valuation: derivation of equation (29). When we elicited the bonus valuation on survey 2, we had not yet told participants whether the bonus would be in effect for period 2 or 3. The theoretical valuations for a period 2 vs. period 3 bonus are identical if we assume that consumers predict no anticipatory effect of the period 3 bonus. Otherwise, this derivation would need to account for the period 2 survey taker's valuation of the perceived internality reduction from the anticipatory effect. Since the actual bonus was for period 3, we focus the derivation on that case and maintain the assumption of zero predicted anticipatory effect.

From the perspective of the period 2 survey taker, the predicted period 3 continuation value (given naivete about future projection bias) as a function of predicted habit stock and period 3 price is

$$V_3(\tilde{s}_3, p_3) = u_3\left(\tilde{x}_3^*(\tilde{s}_3, \tilde{\gamma}, \boldsymbol{p}_3); \tilde{s}_3, p_3\right) + \delta V_4(\tilde{s}_4, \cdot). \tag{153}$$

The change in that predicted continuation value from a marginal change in period 3 price is

$$\frac{dV_3(\tilde{s}_3, p_3)}{dp_3} = \frac{\partial\tilde{u}_3}{\partial p_3} + \frac{\partial\tilde{x}_3}{\partial p_3}\left[\frac{\partial\tilde{u}_3}{\partial\tilde{x}_3} + \delta\frac{dV_4(\tilde{s}_4,\cdot)}{d\tilde{s}_4}\frac{\partial\tilde{s}_4}{\partial\tilde{x}_3}\right]. \tag{154}$$

People taking survey 2 predict that their period 3 selves will set x_3 to maximize that same function with an additional $\tilde{\gamma}x_3$ in period 3 flow utility:

$$\tilde{x}_3^*(\tilde{s}_3, \tilde{\gamma}, \boldsymbol{p}_3) = \arg\max_{x_3} u_3(x_3; \tilde{s}_3, \boldsymbol{p}_3) + \tilde{\gamma}x_3 + \delta V_4(\tilde{s}_4, \cdot). \tag{155}$$

Thus, people taking survey 2 predict that they will set x_3 such that

$$\frac{\partial \tilde{u}_3}{\partial x_3} + \tilde{\gamma} + \delta \frac{dV_4(\tilde{s}_4, \cdot)}{d\tilde{s}_4} \frac{\partial \tilde{s}_4}{\partial \tilde{x}_3} = 0. \tag{156}$$

Substituting equation (156) into equation (154) gives

$$\frac{dV_3(\tilde{s}_3, p_3)}{dp_3} = \frac{\partial \tilde{u}_3}{\partial p_3} - \tilde{\gamma}\frac{\partial \tilde{x}_3}{\partial p_3} \tag{157}$$

$$= -\tilde{x}_3(p_3) - \tilde{\gamma}\frac{\partial \tilde{x}_3}{\partial p_3}. \tag{158}$$

This illustrates a temptation-adjusted envelope theorem: the effect of a marginal price change on the long-run self's utility (given perceived misoptimization from the long-run self's perspective) equals the mechanical effect $\tilde{x}_3(p_3)$ adjusted by the magnitude of the perceived misoptimization $\tilde{\gamma}\frac{\partial \tilde{x}_3}{\partial p_3}$. With zero perceived temptation ($\tilde{\gamma} = 0$), this reduces to the standard envelope theorem. The derivation for a period 2 bonus would be analogous, except with $(1-\alpha)$ multiplying the predicted period 3 continuation value in both the survey taker's objective function and the predicted period 2 objective function.

We integrate over equation (158) to determine the effect of a non-marginal price increase from 0 to p^B:

$$V_3\left(\tilde{s}_3, p_3 = p_3^B\right) - V_3\left(\tilde{s}_3, p_3 = 0\right) = \int_{p_3=0}^{p_3=p_3^B} -\tilde{x}_3(p_3) - \tilde{\gamma}\frac{\partial \tilde{x}_3}{\partial p_3}\, dp_3 \tag{159}$$

$$= -p_3^B \cdot \left(\tilde{x}_3(p_3^B) + \tilde{x}_3(0)\right)/2 - \tilde{\gamma}\cdot\left(\tilde{x}_3(p_3^B) - \tilde{x}_3(0)\right), \tag{160}$$

where the second line follows from the fact that demand is linear in price, which was shown in Proposition 2.

Limit valuation: derivation of equation (31). The period 3 survey-taker's objective function is

$$V_3(s_3, \tilde{\gamma}_3) = u_3\left(x_3^*(s_3, \tilde{\gamma}_3, \boldsymbol{p}_3); s_3, p_3\right) + \begin{array}{c}\alpha \sum_{r=4}^{T} \delta^{r-3} u_r\left(\tilde{x}_r^*(s_3, \tilde{\gamma}, \boldsymbol{p}_r); s_3, p_r\right) \\ +(1-\alpha)\delta V_4(\tilde{s}_4, \cdot)\end{array}. \tag{161}$$

This equation uses the assumption that the survey taker is projection biased.

The change in that objective function from a marginal change in perceived period 3 temptation is

$$\frac{dV_3(s_3, \tilde{\gamma}_3)}{d\tilde{\gamma}_3} = \frac{\partial x_3^*(s_3, \tilde{\gamma}_3, \boldsymbol{p}_3)}{\partial \tilde{\gamma}_3}\left[\frac{\partial u_3}{\partial x_3} + (1-\alpha)\delta\frac{\partial V_4(\tilde{s}_4, \cdot)}{\partial \tilde{s}_4}\frac{\partial \tilde{s}_4}{\partial \tilde{x}_3}\right]. \tag{162}$$

People taking survey 3 predict that they will set x_3^* such that

$$\frac{\partial u_3}{\partial x_3} + \tilde{\gamma}_3 + (1-\alpha)\delta \frac{dV_4(\tilde{s}_4, \cdot)}{d\tilde{s}_4} \frac{\partial \tilde{s}_4}{\partial \tilde{x}_3} = 0. \tag{163}$$

Substituting the period 3 first-order condition from equation (163) into equation (162) gives

$$\frac{dV_3(s_3, \tilde{\gamma}_3)}{d\tilde{\gamma}_3} = -\tilde{\gamma}_3 \frac{\partial x_3^*(s_3, \tilde{\gamma}_3, \boldsymbol{p}_3)}{\partial \tilde{\gamma}_3}. \tag{164}$$

We integrate over equation (164) to determine the effect of the non-marginal temptation reduction from $\tilde{\gamma}$ to $(1-\omega)\tilde{\gamma}$:

$$v^L = V_3(s_3, \tilde{\gamma}_3 = (1-\omega)\tilde{\gamma}) - V_3(s_3, \tilde{\gamma}_3 = \tilde{\gamma}) = \int_{\tilde{\gamma}_3 = \tilde{\gamma}}^{\tilde{\gamma}_3 = (1-\omega)\tilde{\gamma}} -\tilde{\gamma}_3 \frac{\partial x_3^*(\tilde{\gamma}_3)}{\partial \tilde{\gamma}_3} \, d\tilde{\gamma}_3. \tag{165}$$

$$= (x_3^*(\tilde{\gamma}) - x_3^*(0)) \cdot \tilde{\gamma} \cdot \frac{1}{2} - (1-\omega)^2 \cdot (\tilde{x}_3(\tilde{\gamma}) - \tilde{x}_3(0)) \cdot \tilde{\gamma} \cdot \frac{1}{2} \tag{166}$$

$$= (x_3^*(\tilde{\gamma}) - x_3^*(0)) \cdot \tilde{\gamma} \cdot \left(1 - (1-\omega)^2\right) \cdot \frac{1}{2} \tag{167}$$

$$= (x_3^*(\tilde{\gamma}) - x_3^*(0)) \cdot \tilde{\gamma} \cdot \frac{\omega(2-\omega)}{2} \tag{168}$$

$$= \frac{(x_3^*(\tilde{\gamma}) - x_3^*((1-\omega)\tilde{\gamma}))}{\omega} \cdot \tilde{\gamma} \cdot \frac{\omega(2-\omega)}{2} \tag{169}$$

$$= (x_3^*(\tilde{\gamma}) - x_3^*((1-\omega)\tilde{\gamma})) \cdot \tilde{\gamma} \cdot \frac{(2-\omega)}{2}, \tag{170}$$

where the second line is the area of the long-run self's perceived deadweight loss reduction trapezoid (following from linear demand) and the fifth line follows from the assumption that $\tilde{\tau}^L/\omega = \tilde{\tau}^0$.

G.5 Temptation with Multiple Goods

We now extend our model to include a second temptation good y, which in our experiment is FITSBY use on other devices. Habit stock now evolves according to $s_{t+1} = \rho(s_t + x_t + y_t)$. Before period t, consumers now consider flow utility to be $u_t(x_t, y_t; s_t, p_t)$. In period t, consumers choose as if period t flow utility is $u_t(x_t, y_t; s_t, p_t) + \gamma_x x_t + \gamma_y y_t$. Before period t, consumers predict that they will choose as if period t flow utility is $u_t(x_t, y_t; s_t, p_t) + \tilde{\gamma}_x x_t + \tilde{\gamma}_y y_t$. x is still sold at price p_t, while y_t has zero price. The limit treatment fully eliminates perceived and actual temptation on x.

We derive new equations for γ or $\tilde{\gamma}$ for the limit effect, bonus valuation, and limit valuation strategies. With all three strategies, if y is not a temptation good ($\tilde{\gamma}_y = \gamma_y = 0$) or if y is neither a substitute nor a complement for x, then our original estimating equations are unaffected.

Limit effect. To derive γ using the limit effect strategy, we assume full projection bias ($\alpha = 1$). We

assume that the static quadratic flow utility function is now

$$u(x,y;p) = \frac{\eta_x}{2}x^2 + \xi_x x - px + \sigma xy + \frac{\eta_y}{2}y^2 + \xi_y y. \tag{171}$$

Without the limit, consumers maximize $u(x,y;p) + \gamma_x x + \gamma_y y$, giving

$$y^*(x) = \frac{\sigma x + \xi_y + \gamma_y}{-\eta_y} \tag{172}$$

$$x^* = \frac{\xi_x - p + \sigma\frac{\xi_y+\gamma_y}{-\eta_y} + \gamma_x}{-\eta_x + \frac{\sigma^2}{\eta_y}} \tag{173}$$

Taking the expectation over individuals, the bonus effect on x^* is

$$\tau_x^B = \frac{p^B}{-\eta_x + \frac{\sigma^2}{\eta_y}} \tag{174}$$

The limit allows consumers to set x_L before period t. When setting the limit, consumers predict that in period t they will set y conditional on x_L to maximize $u(x_L, y; p) + \tilde{\gamma}_x x_L + \tilde{\gamma}_y y$, giving

$$y^*(x_L) = \frac{\sigma x_L + \xi_y + \tilde{\gamma}_y}{-\eta_y}. \tag{175}$$

Consumers thus set x_L to maximize $u(x_L, y^*(x_L); p)$, giving

$$x_L = \frac{\xi_x - p + \xi_y\frac{\sigma}{-\eta_y}}{-\eta_x + \frac{\sigma^2}{\eta_y}}. \tag{176}$$

The effect of the limit on y is $y^*(x_L) - y^*(x^*) = \frac{\sigma x_L+\xi_y+\tilde{\gamma}_y}{-\eta_y} - \frac{\sigma x^*+\xi_y+\gamma_y}{-\eta_y}$. Taking the expectation over individuals, the limit effect on y is

$$\tau_y^L = \frac{\sigma}{-\eta_y}\tau_x^L \tag{177}$$

The effect of the limit on x is

$$x_L - x^* = \frac{\xi_x - p + \xi_y \frac{\sigma}{-\eta_y}}{-\eta_x + \frac{\sigma^2}{\eta_y}} - \frac{\xi_x - p + \sigma \frac{\xi_y + \gamma_y}{-\eta_y} + \gamma_x}{-\eta_x + \frac{\sigma^2}{\eta_y}} \tag{178}$$

$$= \frac{\frac{\sigma}{-\eta_y}\gamma_y + \gamma_x}{-\eta_x + \frac{\sigma^2}{\eta_y}} \tag{179}$$

$$= \frac{-\gamma\left(1 + \frac{\sigma}{-\eta_y}\right)}{-\eta_x + \frac{\sigma^2}{\eta_y}} \tag{180}$$

where the third line assumes $\gamma_x = \gamma_y = \gamma$.

Taking the expectation over individuals and substituting equations (174) and (177) gives

$$\tau_x^L = \frac{-\gamma\left(1 + \frac{\tau_y^L}{\tau_x^L}\right)}{p^B/\tau_x^B}. \tag{181}$$

Rearranging gives

$$\gamma = \frac{\tau_x^L \cdot (p^B/\tau_x^B)}{1 + \frac{\tau_y^L}{\tau_x^L}}. \tag{182}$$

This exactly parallels equation (28) for the $\alpha = 1$ case, except adjusting the denominator for substitution. If x and y are substitutes, then the estimated γ increases: more temptation is required to explain a given limit when the consumer knows that she can evade the limit through substitution to another temptation good. If x and y are complements, then the estimated γ decreases: less temptation is needed to explain a given limit when the consumer knows that the limit will also cause reductions in another temptation good.

Bonus valuation. The derivation for the bonus valuation with substitute goods is very similar to the one-good case. The change in the period 3 continuation value function from a marginal change in p_3 is

$$\frac{dV_3(\tilde{s}_3, p_3)}{dp_3} = \frac{\partial \tilde{u}_3}{\partial p_3} + \frac{\partial \tilde{x}_3}{\partial p_3}\left[\frac{\partial \tilde{u}_3}{\partial \tilde{x}_3} + \delta \frac{dV_4(\tilde{s}_4, \cdot)}{d\tilde{s}_4}\frac{\partial \tilde{s}_4}{\partial \tilde{x}_3}\right] + \frac{\partial \tilde{y}_3}{\partial p_3}\left[\frac{\partial \tilde{u}_3}{\partial \tilde{y}_3} + \delta \frac{dV_4(\tilde{s}_4, \cdot)}{d\tilde{s}_4}\frac{\partial \tilde{s}_4}{\partial \tilde{y}_3}\right]. \tag{183}$$

People taking survey 2 predict that they will set x_3 and y_3 according to

$$\frac{\partial \tilde{u}_3}{\partial x_3} + \tilde{\gamma}_x + \delta \frac{dV_4(\tilde{s}_4, \cdot)}{d\tilde{s}_4}\frac{\partial \tilde{s}_4}{\partial \tilde{x}_3} = 0 \tag{184}$$

$$\frac{\partial \tilde{u}_3}{\partial y_3} + \tilde{\gamma}_y + \delta \frac{dV_4(\tilde{s}_4, \cdot)}{d\tilde{s}_4}\frac{\partial \tilde{s}_4}{\partial \tilde{y}_3} = 0. \tag{185}$$

Substituting equations (184) and (185) as well as $\frac{\partial \tilde{u}_3}{\partial p_3} = -\tilde{x}_3(p_3)$ into equation (183) gives

$$\frac{dV_3(\tilde{s}_3, p_3)}{dp_3} = -\tilde{x}_3(p_3) - \tilde{\gamma}_x \frac{\partial \tilde{x}_3}{\partial p_3} - \tilde{\gamma}_y \frac{\partial \tilde{y}_3}{\partial p_3}. \tag{186}$$

Integrating over a non-marginal price increase from 0 to p^B assuming linear demand, also assuming $\tilde{\gamma}_x = \tilde{\gamma}_y = \tilde{\gamma}$, taking the expectation over participants, and rearranging gives

$$\tilde{\gamma} = \frac{\bar{v}^B - \bar{F}^B + p_3^B \bar{x}_3^{B+BC}}{-\left(\bar{\tau}_{x3}^B + \bar{\tau}_{y3}^B\right)} \tag{187}$$

This exactly parallels equation (30), except adjusting the denominator for substitution. The survey taker values the total temptation reduction $-\left(\bar{\tau}_{x3}^B + \bar{\tau}_{y3}^B\right)$ induced by the bonus. If x and y are substitutes, the total temptation reduction is lower, and more temptation is needed to justify a given valuation. If x and y are complements, the total temptation reduction is higher, and less temptation is needed to justify a given valuation.

Limit valuation. The derivation for the limit valuation with substitute goods is also similar to the one-good case. The change in the period 3 survey-taker's objective function from a marginal change in perceived period 3 temptation for good x only is

$$\frac{dV_3(s_3, \tilde{\gamma}_{x3})}{d\tilde{\gamma}_{x3}} = \frac{\partial x_3^*}{\partial \tilde{\gamma}_{x3}}\left[\frac{\partial u_3}{\partial x_3} + (1-\alpha)\delta \frac{\partial V_4(\tilde{s}_4, \cdot)}{\partial \tilde{s}_4}\frac{\partial \tilde{s}_4}{\partial \tilde{x}_3}\right] + \frac{\partial y_3^*}{\partial \tilde{\gamma}_{x3}}\left[\frac{\partial u_3}{\partial y_3} + (1-\alpha)\delta \frac{\partial V_4(\tilde{s}_4, \cdot)}{\partial \tilde{s}_4}\frac{\partial \tilde{s}_4}{\partial \tilde{y}_3}\right]. \tag{188}$$

Substituting the predicted period 3 first-order conditions for x and y gives

$$\frac{dV_3(s_3, \tilde{\gamma}_{x3})}{d\tilde{\gamma}_{x3}} = -\tilde{\gamma}_{x3}\frac{\partial x_3^*}{\partial \tilde{\gamma}_{x3}} - \tilde{\gamma}_y \frac{\partial y_3^*}{\partial \tilde{\gamma}_{x3}}. \tag{189}$$

Integrating over this from $\tilde{\gamma}_x$ to $(1-\omega)\tilde{\gamma}_x$ assuming linear demand, also assuming $\tilde{\gamma}_x = \tilde{\gamma}_y = \tilde{\gamma}$, taking the expectation over participants, and rearranging gives

$$\tilde{\gamma} = \frac{\bar{v}^L}{-\left(\bar{\tau}_3^L(2-\omega)/2 + \bar{\tau}_{y3}^L\right)}. \tag{190}$$

As with the bonus valuation, the survey taker values the total temptation deadweight loss reduction induced by the limit. If x and y are substitutes, the total temptation reduction is lower, and more temptation is needed to justify a given valuation. If x and y are complements, the total temptation reduction is higher, and less temptation is needed to justify a given valuation.

G.6 Intercept

Derivation of equation (33).

Re-arranging steady state consumption from equation (21) gives

$$(1-\alpha)\delta\rho(\phi-\xi)+\xi-(1-(1-\alpha)\delta\rho)\,p+(1-\alpha)\delta\rho\left[(\zeta-\eta)m_{ss}-\left(1+\bar{\lambda}\right)\bar{\gamma}\right]+\gamma=$$
$$x_{ss}\left[-\eta-(1-\alpha)\delta\rho(\zeta-\eta)-\zeta\frac{\rho-(1-\alpha)\delta\rho^2}{1-\rho}\right]. \qquad (191)$$

Solving for the intercept and substituting $x_{i1}=x_{ss}$ gives equation (33).

H Counterfactual Simulations Appendix

Table A14: **Effects of Temptation and Habit Formation on FITSBY Use**

FITSBY use (minutes/day)	(1) Restricted model ($\tau_2^B=0$, $\alpha=1$)	(2) Unrestricted model ($\alpha=\hat{\alpha}$)
Baseline	153	153
	[149, 157]	[149, 157]
No naivete	153	151
	[149, 157]	[140, 156]
No temptation	105	103
	[76.9, 120]	[67.0, 119]
No habit formation	78.1	73.3
	[50.2, 102]	[43.1, 99.4]
No temptation or habit formation	53.8	49.0
	[25.6, 77.9]	[17.2, 75.7]

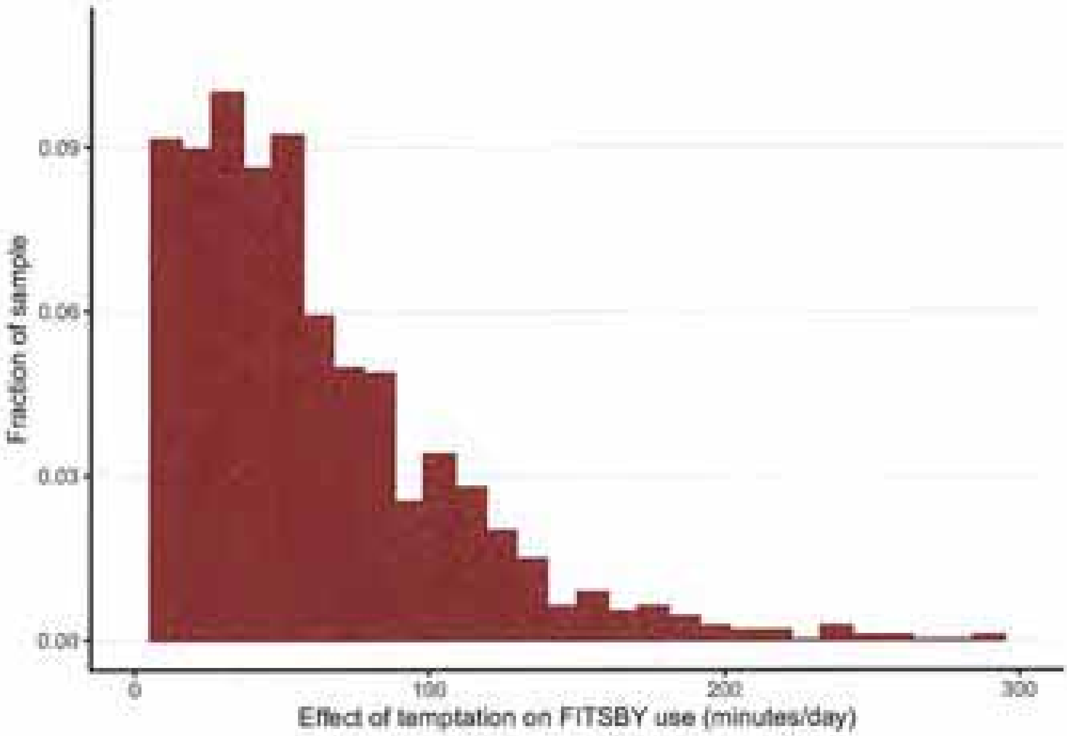
Notes: This table presents point estimates and bootstrapped 95 percent confidence intervals for predicted steady-state FITSBY use with different parameter assumptions, using equation (15). The numbers are as plotted in Figure 10.

Table A15: **Effects of Temptation on FITSBY Use Under Alternative Assumptions**

Effect of temptation on FITSBY use (minutes/day)	(1) Restricted model ($\tau_2^B = 0$, $\alpha = 1$)	(2) Unrestricted model ($\alpha = \hat{\alpha}$)
Limit effect	47.5	49.5
	[34.3, 75.0]	[34.9, 86.4]
Bonus valuation	70.5	71.2
	[49.2, 116]	[49.5, 118]
Limit valuation	61.5	62.3
	[42.8, 103]	[43.3, 106]
Limit effect, multiple-good model	57.4	
	[40.0, 97.0]	
Bonus valuation, multiple-good model	63.2	63.9
	[44.5, 103]	[44.7, 107]
Limit valuation, multiple-good model	91.3	91.6
	[50.1, 155]	[51.0, 155]
Limit effect, $\omega = \hat{\omega}$	123	127
	[85.2, 155]	[87.3, 156]
Limit valuation, $\omega = \hat{\omega}$	42.7	43.8
	[29.7, 71.1]	[30.2, 76.6]
Heterogeneous limit effect	47.1	48.6
	[34.2, 71.9]	[34.6, 76.5]
Limit effect, weighted sample	52.2	57.8
	[32.3, 112]	[33.9, 144]

Notes: This table presents point estimates and bootstrapped 95 percent confidence intervals for the effects of temptation on average steady-state FITSBY use, using equation (15). The first nine estimates are for the nine temptation estimation strategies presented in Table A9. The tenth estimate is for the limit effect strategy after reweighting the sample to be more representative of U.S. adults. Appendix Tables A11–A13 present the demographics, moments, and parameter estimates in the weighted sample. Average baseline FITSBY use is 153 and 156 minutes per day for the unweighted and weighted samples, respectively. We do not have a limit effect estimate for the unrestricted multiple-good model.

Figure A35: **Distribution of Effects of Temptation on FITSBY Use**



Notes: Using the heterogeneous limit effect strategy, we estimate temptation $\hat{\gamma}_i$ for each Limit group participant, which we then insert into equation (15) to predict the individual-specific effect of temptation on steady-state FITSBY use. This figure presents the distribution of effects across participants, winsorized at 300 minutes per day.

JAMA Pediatrics | Original Investigation

Association of Habitual Checking Behaviors on Social Media With Longitudinal Functional Brain Development

Maria T. Maza, BS; Kara A. Fox, MA; Seh-Joo Kwon, BS; Jessica E. Flannery, PhD; Kristen A. Lindquist, PhD; Mitchell J. Prinstein, PhD; Eva H. Telzer, PhD

Supplemental content

IMPORTANCE Social media platforms provide adolescents with unprecedented opportunities for social interactions during a critical developmental period when the brain is especially sensitive to social feedback.

OBJECTIVE To explore how adolescents' frequency of checking behaviors on social media platforms is associated with longitudinal changes in functional brain development across adolescence.

DESIGN, SETTING, AND PARTICIPANTS A 3-year longitudinal cohort study of functional magnetic resonance imaging (fMRI) among sixth- and seventh-grade students recruited from 3 public middle schools in rural North Carolina.

EXPOSURES At wave 1, participants reported the frequency at which they checked Facebook, Instagram, and Snapchat.

MAIN OUTCOME OR MEASURE Neural responses to the Social Incentive Delay task when anticipating receiving social feedback, measured annually using fMRI for 3 years. Participants saw a cue that indicated whether the social feedback (adolescent faces with emotional expressions) would be a reward, punishment, or neutral; after a delay, a target appeared and students responded by pressing a button as quickly as possible; a display of social feedback depended on trial type and reaction time.

RESULTS Of 178 participants recruited at age 12 years, 169 participants (mean [SD] age, 12.89 [0.58] years; range, 11.93-14.52 years; 91 [53.8%] female, 38 [22.5%] Black, 60 [35.5%] Latinx, 50 [29.6%] White, 15 [8.9%] multiracial) met the inclusion criteria. Participants with habitual social media checking behaviors showed lower neural sensitivity to social anticipation at age 12 years compared with those with nonhabitual checking behaviors in the left amygdala, posterior insula (PI), and ventral striatum (VS; β, −0.22; 95% CI, −0.33 to −0.11), right amygdala (β, −0.19; 95% CI, −0.30 to −0.08), right anterior insula (AI; β, −0.23; 95% CI, −0.37 to −0.09), and left dorsolateral prefrontal cortex (DLPFC; β, −0.29; 95% CI, −0.44 to −0.14). Among those with habitual checking behaviors, there were longitudinal increases in the left amygdala/PI/VS (β, 0.11; 95% CI, 0.04 to 0.18), right amygdala (β, 0.09; 95% CI, 0.02 to 0.16), right AI (β, 0.15; 95% CI, 0.02 to 0.20), and left DLPFC (β, 0.19; 95% CI, 0.05 to 0.25) during social anticipation, whereas among those with
behaviors, longitudinal decreases were seen in the left amygdala/PI
−0.19 to −0.06), right amygdala (β, −0.10; 95% CI, −0.17 to −0.03),
−0.22 to −0.04), and left DLPFC (β, −0.10; 95% CI, −0.22 to −0.03).

CONCLUSIONS AND RELEVANCE The results of this cohort study sugg
checking behaviors in early adolescence may be associated with ch
sensitivity to social rewards and punishments. Further research exa
associations between social media use, adolescent neural developm
adjustment is needed to understand the effects of a ubiquitous infl
today's adolescents.

No Annotations

JAMA Pediatr. 2023;177(2):160-167. doi:10.1001/jamapediatrics.2022.4924
Published online January 3, 2023. Corrected on February 13, 2023.

JAMA Pediatrics | Original Investigation

Association of Habitual Checking Behaviors on Social Media With Longitudinal Functional Brain Development

Maria T. Maza, BS; Kara A. Fox, MA; Seh-Joo Kwon, BS; Jessica E. Flannery, PhD; Kristen A. Lindquist, PhD; Mitchell J. Prinstein, PhD; Eva H. Telzer, PhD

Supplemental content

IMPORTANCE Social media platforms provide adolescents with unprecedented opportunities for social interactions during a critical developmental period when the brain is especially sensitive to social feedback.

OBJECTIVE To explore how adolescents' frequency of checking behaviors on social media platforms is associated with longitudinal changes in functional brain development across adolescence.

DESIGN, SETTING, AND PARTICIPANTS A 3-year longitudinal cohort study of functional magnetic resonance imaging (fMRI) among sixth- and seventh-grade students recruited from 3 public middle schools in rural North Carolina.

EXPOSURES At wave 1, participants reported the frequency at which they checked Facebook, Instagram, and Snapchat.

MAIN OUTCOME OR MEASURE Neural responses to the Social Incentive Delay task when anticipating receiving social feedback, measured annually using fMRI for 3 years. Participants saw a cue that indicated whether the social feedback (adolescent faces with emotional expressions) would be a reward, punishment, or neutral; after a delay, a target appeared and students responded by pressing a button as quickly as possible; a display of social feedback depended on trial type and reaction time.

RESULTS Of 178 participants recruited at age 12 years, 169 participants (mean [SD] age, 12.89 [0.58] years; range, 11.93-14.52 years; 91 [53.8%] female; 38 [22.5%] Black; 60 [35.5%] Latinx; 50 [29.6%] White; 15 [8.9%] multiracial) met the inclusion criteria. Participants with habitual social media checking behaviors showed lower neural sensitivity to social anticipation at age 12 years compared with those with nonhabitual checking behaviors in the left amygdala, posterior insula (PI), and ventral striatum (VS; β, −0.22; 95% CI, −0.33 to −0.11), right amygdala (β, −0.19; 95% CI, −0.30 to −0.08), right anterior insula (AI; β, −0.23; 95% CI, −0.37 to −0.09), and left dorsolateral prefrontal cortex (DLPFC; β, −0.29; 95% CI, −0.44 to −0.14). Among those with habitual checking behaviors, there were longitudinal increases in the left amygdala/PI/VS (β, 0.11; 95% CI, 0.04 to 0.18), right amygdala (β, 0.09; 95% CI, 0.02 to 0.16), right AI (β, 0.15; 95% CI, 0.02 to 0.20), and left DLPFC (β, 0.19; 95% CI, 0.05 to 0.25) during social anticipation, whereas among those with nonhabitual checking behaviors, longitudinal decreases were seen in the left amygdala/PI/VS (β, −0.12; 95% CI, −0.19 to −0.06), right amygdala (β, −0.10; 95% CI, −0.17 to −0.03), right AI (β, −0.13; 95% CI, −0.22 to −0.04), and left DLPFC (β, −0.10; 95% CI, −0.22 to −0.03).

CONCLUSIONS AND RELEVANCE The results of this cohort study suggest that social media checking behaviors in early adolescence may be associated with changes in the brain's sensitivity to social rewards and punishments. Further research examining long-term associations between social media use, adolescent neural development, and psychological adjustment is needed to understand the effects of a ubiquitous influence on development for today's adolescents.

Author Affiliations: Department of Psychology and Neuroscience, University of North Carolina at Chapel Hill, Chapel Hill.

Corresponding Author: Eva H. Telzer, PhD, 235 E Cameron Ave, Chapel Hill, NC 27599 (ehtelzer@unc.edu).

JAMA Pediatr. 2023;177(2):160-167. doi:10.1001/jamapediatrics.2022.4924
Published online January 3, 2023. Corrected on February 13, 2023.

In the span of a generation, social media has dramatically changed the landscape of adolescent development, providing unprecedented opportunities for social interactions around the clock.[1] Social media provides a constant and unpredictable stream of social inputs to adolescents during a critical developmental period when the brain becomes especially sensitive to social rewards and punishments.[2] Motivated by the anticipation of this social feedback, adolescents' constant, habitual checking of social media may alter neurodevelopment, significantly changing the ways in which the adolescent brain responds to its environment.

Social media allows immediate access to social information at any time it is desired[1,3,4] and is designed to hold users' engagement by maximizing social rewards. "Likes," notifications, and messages arrive unpredictably on a maximally powerful variable reinforcement schedule, conditioning individuals to check social media habitually in anticipation of this social feedback.[5] With 78% of 13- to 17-year-olds reporting checking their devices at least hourly[6] and 46% checking "almost constantly,"[7] adolescents may be uniquely vulnerable to habitual checking behaviors.

The brain undergoes significant structural and functional reorganization during adolescence.[8] Neural regions involved in motivational relevance (eg, the ventral striatum; VS) and affective salience (eg, the amygdala and insula) become hyperactive, orienting teens to rewarding stimuli in their environment, particularly from peers.[9-14] Adolescents' habitual checking of social media may be exacerbating an already enhanced neural response to the anticipation of salient social feedback. Additionally, the motivational salience of social contexts may undermine adolescents' ability to engage in cognitive control and, subsequently, to regulate their behaviors.[15] Consequently, repeated exposure to digital social rewards (eg, notifications or likes) may increase neural reactivity to reward-related cues, reducing adolescents' ability to resist urges to check social media.[16,17]

The current study aimed to examine whether social media use is associated with longitudinal changes in functional brain development across adolescence, a developmental period characterized by peak social media use[18] and heightened neural sensitivity to social feedback from peers.[8] We hypothesized that checking social media habitually would make adolescents increasingly hypersensitive to social feedback anticipation and thus would be associated with longitudinal increases in neural activation, particularly within regions comprising the motivational (eg, VS), affective salience (eg, insula and amygdala), and cognitive control (eg, dorsolateral prefrontal cortex; DLPFC) networks. Conversely, we hypothesized that nonhabitual checking would be associated with longitudinal decreases in neural activation in the same brain regions. Given the limited research exploring longitudinal neural activation in relation to social media behaviors, we conducted exploratory whole-brain analyses to determine which brain regions showed the greatest differences in neural activation longitudinally. To our knowledge, results from this study would provide the first insight into how habitual social media behaviors may be altering adolescent brain development.

Key Points

Question Is adolescents' frequency of checking behaviors on 3 social media platforms (Facebook, Instagram, Snapchat) associated with longitudinal changes in functional brain development across adolescence?

Findings In this cohort study of 169 sixth- and seventh-grade students, participants who engaged in habitual checking behaviors showed a distinct neurodevelopmental trajectory within regions of the brain comprising the affective salience, motivational, and cognitive control networks in response to anticipating social rewards and punishments compared with those who engaged in nonhabitual checking behaviors.

Meaning These results suggest that habitual checking of social media in early adolescence may be longitudinally associated with changes in neural sensitivity to anticipation of social rewards and punishments, which could have implications for psychological adjustment.

Methods

Participants

Participants were recruited from a larger, school-based study of 873 sixth- and seventh-grade students from 3 public rural middle schools in North Carolina to participate in a longitudinal functional magnetic resonance imaging (fMRI) study. We recruited 2 cohorts of participants at 12 to 13 years of age across 2 years of the study, leading to a sample size of 178 adolescents (148 students for cohort 1 and 30 for cohort 2). Of the recruited participants for cohort 1, 5 met exclusion criteria after consenting to the study and thus were excluded and not invited back for later waves (see the eMethods in the Supplement for exclusion criteria). Across all waves, 25 participants completed 1 time point, 36 completed 2 time points, and 112 completed 3 time points. All participants provided written informed consent or assent, and the University's Institutional Review Board approved all aspects of the study. Race and ethnicity were self-reported by participants. This study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline. For more information on study procedures, see the eMethods in the Supplement.

Self-reported Social Media Use

Participants reported frequency of checking at wave 1 only. For 3 popular social media platforms (Facebook, Instagram, and Snapchat), participants were asked how many times per day they checked each platform, with answers grouped into 8 numerical score categories (1, <1 time per day; 2, 1 time per day; 3, 2-3 times per day; 4, 4-5 times per day; 5, 6-10 times per day; 6, 11-15 times per day; 7, 16-20 times per day; 8, >20 times per day). We recoded participants' scores to create an ordinal scale that captured social media checking frequency across a meaningful distribution that could be assessed quantitatively. A score of 1 was recoded to 0 and a score of 2 was recoded to 1. Scores between 3 and 7 were recoded to the average of the range of

number of times checked; for example, if participants selected 6 for their Facebook use (ie, checked Facebook between 11 and 15 times per day), then their score was recoded to the average of 11 and 15 times, which in this case was 13 times checked. Reported scores of 8 (ie, checked >20 times per day) were recoded to 20 times checked. For each participant, the recoded checking behaviors on the 3 social media platforms were summed to create a total social media checking score that ranged from 0 to 54 (mean [SD] checking behaviors per day, 11.85 [15.39]).

Social Incentive Delay Task

At each wave, participants attended a brain imaging session during which they completed the Social Incentive Delay task while undergoing fMRI to measure neural responses when anticipating receiving social rewards and avoiding social punishments.[19,20] On each trial, participants saw a cue (for 500 milliseconds) indicating whether the potential social feedback would be a reward, punishment, or neutral. After a variable delay (mean delay, 2000 milliseconds; range, 480-3900 milliseconds), a target appeared (for 300 milliseconds), at which point participants were instructed to respond by pressing a button as quickly as possible. The display of social feedback (for 1450 milliseconds) depended on the trial type and participants' reaction time. In the social reward condition, happy faces were the outcome of a fast response (hit), and blurred faces were the outcome of a slow response (miss). In the social punishment condition, a hit earned a blurred face, and a miss earned an angry face. In the control condition, a blurred face was always the outcome for both hits and misses. Trials were presented in an event-related design, with reward, punishment, and neutral trials randomly ordered. Participants completed 2 rounds of the task, totaling 116 trials (48 reward, 48 punishment, and 20 neutral trials).

Task difficulty was standardized to a hit rate of approximately 50% for all participants by adjusting target duration to individual reaction times. Age-matched adolescent faces with emotional expressions of 24 ethnically diverse people (12 female) were used as reward and punishment stimuli. Photographs were taken from the National Institute of Mental Health Child Emotional Faces Picture Set. Participants were trained on the meaning of each cue and completed 12 practice trials prior to entering the scanner.

Statistical Analysis

fMRI Data Acquisition

Imaging data were collected using a 3-T Magnatom Prisma MRI scanner (Siemens Healthineers). For specific fMRI image acquisition parameters and preprocessing methods, see the eMethods in the Supplement. Individual level, fixed-effects analyses were estimated using the general linear model convolved with a canonical hemodynamic response function in Statistical Parametric Mapping software package SPM12 (Wellcome Centre for Human Neuroimaging, UCL Queen Square Institute of Neurology). The task was modeled as event related with 8 conditions, including 3 anticipation conditions (reward, punishment, and neutral), 2 outcome conditions for both reward (hit or miss) and punishment (hit or miss), and 1 outcome condition for neutral. Anticipation conditions were modeled as the onset of the cue and a duration of zero, and outcome conditions were modeled at the onset of the outcome with a duration of zero. Six motion parameters were modeled as nuisance regressors. Using the general linear model, linear contrast images comparing each of the conditions of interest were calculated for each individual. The primary contrasts of interest for this study were reward anticipation vs neutral anticipation and punishment anticipation vs neutral anticipation, given our supposition that checking behaviors on social media platforms is motivated by the anticipation of social feedback.

Longitudinal Whole-Brain Analyses

We conducted longitudinal whole-brain analyses using the 3dLMEr program (AFNI).[21] This program allows for voxel-level whole-brain analysis of linear mixed effects (maximum likelihood, multilevel model). Missing data across waves were accounted for by using full information maximum likelihood, which provides an estimate of the value of a population parameter most likely to result in the observed data even in the presence of missing data.[22] We modeled a 3-way interaction with age (minimum centered), condition (reward and punishment anticipation), and social media checking to assess whether age-related changes in neural activation during social anticipation differed as a function of the type of social anticipation (ie, reward vs punishment) and amount of social media checking behaviors. To correct for multiple comparisons, we conducted a Monte Carlo simulation using the 3dFWHMx and 3dClustSim programs (AFNI)[23] and the group-level brain mask. Smoothness was estimated with the -acf option (-acf a, b, and c parameters 0.55, 4.61, 12.32), which used an average of individual-level autocorrelation function parameters (obtained using each participant's residuals from the first-level model). This simulation indicated that a $P < .05$ family-wise error corrected would be achieved with a voxelwise threshold of $P < .001$ and a minimum cluster size of 80 voxels. A 2-sided $P < .01$ indicated statistical significance.

To explore any significant whole-brain interactions and plot the trajectories, we extracted parameter estimates from significant clusters. Parameter estimates were fitted into a conditional linear trajectory model whereby these post hoc analyses allowed us to unpack the significant 3-way interaction between age, condition, and social media checking behavior. For plotting purposes, we categorized the total social media checking scores as high (>15; habitual), moderate (1-15), and low (0; nonhabitual). This allowed us to test whether trajectories of neural response differed as a function of anticipation type and amount of social media checking.

Results

After exclusions, the final sample size was 169 (mean [SD] age, 12.89 [0.58] years; range, 11.93-14.52 years; 91 [53.8%] female; 38 [22.5%] Black, 60 [35.5%] Latinx, 50 [29.6%] White, 15 [8.9%] multiracial [2 or more racial categories identified other Hispanic or Latinx], and 6 [3.6%] categorized as other [American Indian or Alaska Native, Asian, and Native Hawaiian

or other Pacific Islander]) collected across 3 waves; 136 participants completed wave 1 (mean [SD] age, 12.80 [0.52] years; range, 11.9-14.5 years; 71 [52.2%] female), 131 participants completed wave 2 (mean [SD] age, 13.7 [0.59] years; range, 12.4-15.4 years; 68 [51.9%] female), and 124 participants completed wave 3 (mean [SD] age, 14.70 [0.60] years; range, 13.4-16.3 years; 61 [49.2%] female). The mean (SD) time between waves 1 and 2 was 49.8 (3.9) weeks, and that between waves 2 and 3 was 52.9 (6.9) weeks. Retention was 81.1% from waves 1 to 2 and 85.3% from waves 2 to 3. Adolescents reported checking behaviors on 3 social media platforms at wave 1 only. For descriptive statistics regarding checking behaviors on all 3 platforms, see the eFigure in the Supplement. Checking behaviors within the 3 apps were recoded and summed for a total social media checking score, which ranged from 0 to 54 mean (SD) score, 11.85 (15.39).

Using 3dLMEr to model longitudinal whole-brain changes in sensitivity to social anticipation, there was not a 3-way interaction between type of social anticipation, age, and social media checking behavior, so we collapsed social reward and social punishment. We found significant 2-way interactions between age and social media checking behaviors in several regions, including the posterior insula (PI; x, 34; y, 6; z, −4), the left amygdala (x, −26; y, −2; z, −12), the VS (x, −24; y, 14; z, −4), the right amygdala (x, 22; y, 4; z, −18), anterior insula (AI; x, 36; y, 22; z, −4), and the DLPFC (x, 42; y, −42; z, 28) (**Table**). Of particular interest were the left amygdala extending into the PI and VS (**Figure 1A**), the right amygdala (**Figure 2A**), right AI (**Figure 3A**), and left DLPFC (**Figure 4A**). Significant 2-way interactions between age and social media checking behaviors were found in similar brain regions when receiving social feedback (eTable in the Supplement).

We extracted parameter estimates from each participant at each wave from the significant clusters in order to unpack the 2-way interaction. We ran post hoc conditional linear growth models to compare the trajectories of adolescents who engaged in low (nonhabitual; n = 79), moderate (n = 34), or high (habitual; n = 56) social media checking behaviors. Participants with high (habitual) checking behaviors showed a lower neural sensitivity to social anticipation at age 12 years (ie, the intercept) compared with those with low (nonhabitual) checking behaviors in the left amygdala/PI/VS (β, −0.22; 95% CI, −0.33 to −0.11 [Figure 1B]), right amygdala (β, −0.19; 95% CI, −0.30 to −0.08 [Figure 2B]), right AI (β, −0.23; 95% CI, −0.37 to −0.09 [Figure 3B]), and left DLPFC (β, −0.29; 95% CI, −0.44 to −0.14 [Figure 4B]). Here, β values refer to the main result of checking behavior at age 12 years where negative values indicate lower neural activation with higher checking behavior.

Developmentally, participants with high checking behaviors at age 12 years showed longitudinal increases (ie, the linear slope) in neural sensitivity in the left amygdala/PI/VS (β, 0.11; 95% CI, 0.04 to 0.18 [Figure 1B]), right amygdala (β, 0.09; 95% CI, 0.02 to 0.16 [Figure 2B]), right AI (β, 0.15; 95% CI, 0.02 to 0.20 [Figure 3B]), and left DLPFC (β, 0.19; 95% CI, 0.05 to 0.25 [Figure 4B]). Participants with low checking behaviors showed significant longitudinal decreases in neural sensitivity in the left amygdala/PI/VS (β, −0.12; 95% CI, −0.19 to −0.06 [Figure 1B]), right amygdala (β, −0.10; 95% CI, −0.17 to −0.03 [Figure 2B]), right AI (β, −0.13; 95% CI, −0.22 to −0.04 [Figure 3B]), and small decreases in the left DLPFC (β, −0.10, 95% CI, −0.22 to −0.03 [Figure 4B]). Here, β represents the age-related change in neural activation for each group. Results suggest that trajectories of neural sensitivity to anticipation of social feedback for habitual and nonhabitual checkers are inversely related.

Table. Age-Related Neural Changes as a Function of Social Media Checking During Anticipation of Social Feedback

Anatomical region	MNI coordinates[a] x	y	z	t Statistic	Cluster size, voxels[b]
Posterior insula	34	6	−4	46.1	2038
Left amygdala	−26	−2	−12	46.1	2038
Ventral striatum	−24	14	−4	46.1	2038
Orbitofrontal cortex	−20	−6	−18	38.3	1027
Right amygdala	22	4	−18	38.2	1027
Cerebellum	0	80	−32	30.3	736
Thalamus	12	26	0	35.2	640
Frontal operculum	−56	−14	−8	31.8	606
Anterior insula	36	22	−4	31.8	606
Middle cingulate cortex	−18	38	38	41.4	534
Dorsolateral prefrontal cortex	42	−42	28	44.1	501
Globus pallidus	34	2	66	27.1	456
Inferior temporal gyrus	−58	38	−22	45.8	374
Postcentral gyrus	54	18	28	31.1	309
Hippocampus	−32	8	−32	26	235
Somatosensory area	−60	8	38	26.3	179
Superior temporal gyrus	60	32	12	26.3	177
Supramarginal gyrus	44	36	28	25.2	153
Cerebellar vermis	−4	50	−6	26.8	142
Intraparietal sulcus	−22	88	16	19.1	116
Supplementary motor area	−8	−4	44	20.7	114
Inferior parietal lobule	−52	42	56	27.8	106
Cuneus	12	88	34	23.4	102
Anterior inferior parietal lobule	−50	34	38	23.9	102
Intraparietal sulcus	−28	58	50	20.5	83

Abbreviation: MNI, Montreal Neurological Institute.

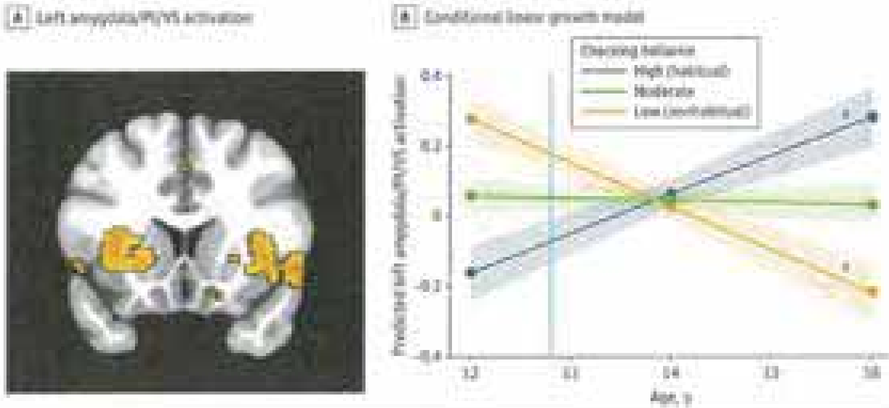
[a] Values are the MNI coordinates to regions of the brain that changed significantly over age.

[b] Clusters that survived cluster extent threshold correction when modeling longitudinal whole-brain changes in sensitivity to social anticipation as a function of social media checking behaviors, assessed using the 3dLMEr program (AFNI). Multiple brain regions may lie within the same brain cluster.

Discussion

This cohort study examined whether early adolescents' frequency of checking behaviors on 3 popular social media platforms (Facebook, Instagram, and Snapchat) was associated with trajectories of functional brain development across adolescence. Adolescents who engaged in high (habitual) checking behaviors showed distinct neural trajectories when anticipating social feedback compared with those who engaged in

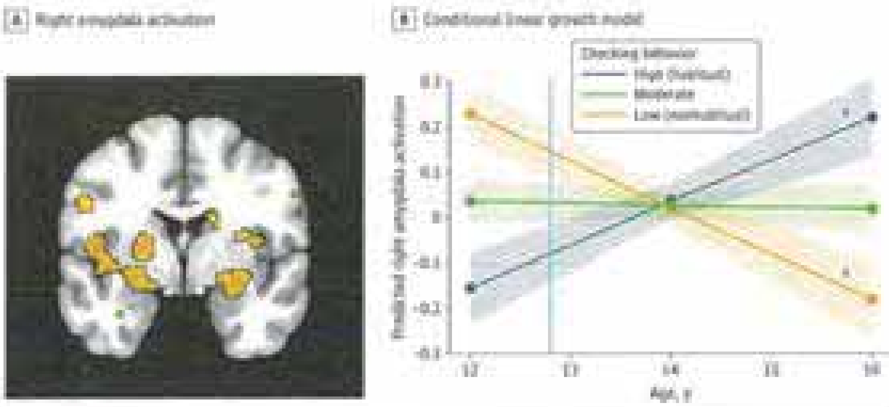
Figure 1. Functional Activation in the Left Amygdala, Posterior Insula (PI), and Ventral Striatum (VS) During the Anticipation of Social Feedback



A, Left amygdala (x, −26; y, −2; z, −12)/PI (x, 34; y, 6; z, −4)/VS (x, −24; y, 14; z, −4) activation during social anticipation differed significantly across time, as a function of social media checking behaviors. B, The dotted vertical line indicates the mean age of participants when they reported the number of checks at wave 1. The shaded areas represent the spread of the data based on the SE of each group.

* $P < .05$.

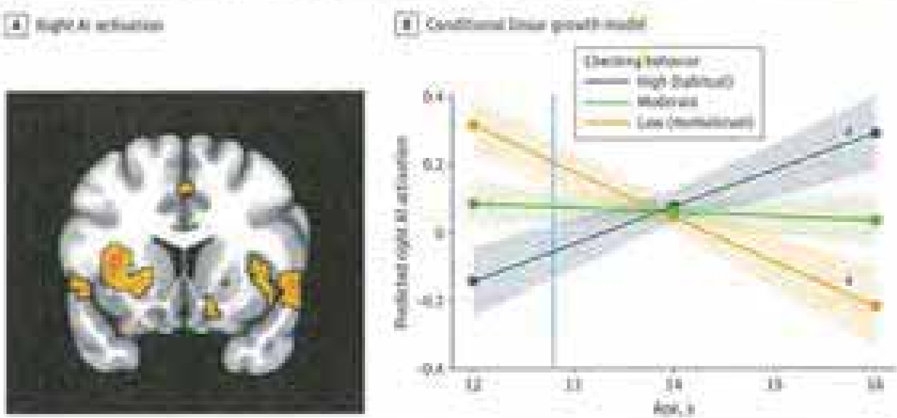
Figure 2. Functional Activation in the Right Amygdala During the Anticipation of Social Feedback



A, Right amygdala (x, 22; y, 4; z, −18) activation during social anticipation significantly differed across time as a function of social media checking behaviors. B, The dotted vertical line indicates the mean age of participants when they reported the number of checks at wave 1. The shaded areas represent the spread of the data based on the SE of each group.

* $P < .05$.

Figure 3. Functional Activation in the Right Anterior Insula (AI) During the Anticipation of Social Feedback



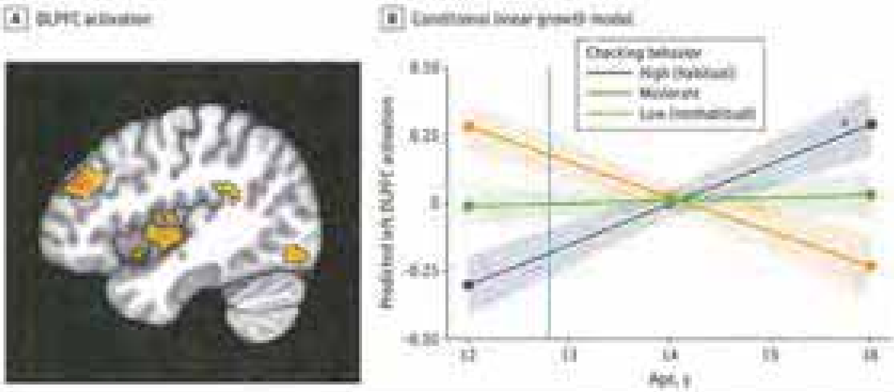
A, Right AI (x, 36; y, 22; z, −4) activation during social anticipation significantly differed across time as a function of social media checking behaviors. B, The dotted vertical line indicates the mean age of participants when they reported the number of checks at wave 1. The shaded areas represent the spread of the data based on the SE of each group.

* $P < .05$.

moderate or low (nonhabitual) checking behaviors, suggesting that habitual social media checking early in adolescence is associated with divergent brain development over time.

We found that 12-year-old adolescents showed different neural patterns based on their social media checking behavior. While participants with habitual checking behaviors demonstrated hypoactivation of the amygdala, PI, VS, and DLPFC in response to anticipation of social feedback, those with nonhabitual behaviors demonstrated hyperactivation in these same brain regions. Interestingly, these patterns diverged

Figure 4. Functional Activation in the Left Dorsolateral Prefrontal Cortex (DLPFC) During the Anticipation of Social Feedback



A, DLPFC (x, 42; y, −42; z, 20) activation during social anticipation significantly differed across time, as a function of social media checking behaviors. B, The dotted vertical line indicates the mean age of participants when they reported the number of checks at wave 1. The shaded areas represent the spread of the data based on the SE of each group.

*P = .05.

across development, with those with habitual behaviors showing longitudinal increases in activation in these regions and those with nonhabitual behaviors showing longitudinal decreases in activation.

Longitudinal decreases in neural activation among participants with nonhabitual checking behaviors may indicate a developmentally normative decreasing sensitivity to social anticipation. Indeed, prior research[12,24-27] has found that in response to social anticipation, early adolescents show an initial hypersensitivity, followed by a decrease in activation of the PI and VS, brain regions associated with salience and motivation, respectively. Additionally, activation of the DLPFC during inhibitory control normatively decreases across adolescence.[28] Decreasing DLPFC activation observed among nonhabitual checkers may indicate that these adolescents are better able to control impulsive or habitual behaviors, such as checking social media, and thus recruit prefrontal cortical regions less over time. In contrast, those with habitual checking behaviors showed longitudinal increases in neural activation in the amygdala, VS, PI, and DLPFC. Research has shown that with constant reinforcement, dopaminergic neurons within salience-related brain regions (ie, the VS, PI, and amygdala) become increasingly responsive to social feedback,[29] and the enhanced value of rewards in the salience and motivation networks may override inhibitory control exerted by the PFC and cause a positive-feedback loop.[30] The observed increase in DLPFC activation may indicate that more effort is required for cognitive control when anticipating social feedback.

Our findings suggest that checking behaviors on social media in early adolescence may tune the brain's sensitivity to potential social rewards and punishments. Whereas individuals with habitual checking behaviors showed initial hypoactivation but increasing sensitivity to potential social cues over time, those with nonhabitual checking behaviors showed initial hyperactivation and decreasing sensitivity over time. Two primary theories contend over whether hypo- or hyperresponsivity to rewards is more associated with behavior.[31] The hyperresponsive theory posits that adolescent reward-associated behaviors are associated with greater activation of the ventral-striatal dopamine circuit.[11,24,32] Consequently, adolescents would experience an increased dopaminergic release in response to social feedback and rewards, which further encourages high-reward behaviors. Indeed, compared with children and adults, adolescents show higher activation in the reward system when receiving rewards.[10-12,28,29] In contrast, the hyporesponsive theory posits that adolescent reward-seeking behaviors may be associated with a deficit in the activity of brain regions associated with motivation.[31,33] This theory argues that repeated exposure to a social reward downregulates dopamine receptors and production, which results in decreased sensitivity of reward circuits. Studies[34,35] suggest that, as adolescents experience fewer or less intense positive feelings from previously rewarding stimuli, they are driven to pursue new appetitive reinforcements through increases in reward-seeking behaviors, which increases activity in dopamine-related circuitry. Indeed, relative to adults, adolescents show less engagement of the VS in anticipation of rewards.[35,36] While for some individuals with habitual checking behaviors, an initial hyposensitivity to potential social rewards and punishments followed by hypersensitivity may contribute to checking behaviors on social media becoming compulsive and problematic, for others, this change in sensitivity may reflect an adaptive behavior that allows them to better navigate their increasingly digital environment.

Limitations

This study has limitations. Notably, because differences in neural trajectories already existed between participants with habitual and nonhabitual checking behaviors at the start of the study, it is difficult to determine whether social media use prior to data collection caused these distinct neural trajectories or preexisting differences in neural activation placed some youth at risk for more habitual checking behaviors. Future studies should explore the neurodevelopmental trajectories of social feedback responsiveness from an earlier age to uncover causal pathways behind this association. Moreover, examination of social media checking behaviors across time is needed to further elucidate associations with development. Finally, future work should examine functional connectivity to explore how

affective salience, motivational, and cognitive control networks coactivate and function at a network level.

Conclusions

Adolescent social media use has proliferated extensively in the past decade. This longitudinal cohort study suggests that social media behaviors in early adolescence may be associated with changes in adolescents' neural development, specifically neural sensitivity to potential social feedback. Further research examining long-term prospective associations between social media use, adolescent neural development, and psychological adjustment is needed to understand the effects of a ubiquitous influence on development for today's adolescents.

ARTICLE INFORMATION

Accepted for Publication: September 29, 2022.

Published Online: January 3, 2023. doi:10.1001/jamapediatrics.2022.4924

Correction: This article was corrected on February 13, 2023, to fix errors in Figures 1 through 4.

Author Contributions: Ms Maza and Dr Telzer had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Mss Maza and Fox contributed equally to this work and are designated as co-first authors.

Concept and design: Maza, Fox, Flannery, Lindquist, Prinstein, Telzer.

Acquisition, analysis, or interpretation of data: All authors.

Drafting of the manuscript: Maza, Fox, Kwon, Lindquist, Prinstein, Telzer.

Critical revision of the manuscript for important intellectual content: Maza, Fox, Flannery, Lindquist, Prinstein, Telzer.

Statistical analysis: Maza, Kwon, Flannery, Telzer.

Obtained funding: Lindquist, Prinstein, Telzer.

Administrative, technical, or material support: Prinstein, Telzer.

Supervision: Lindquist, Prinstein, Telzer.

Conflict of Interest Disclosures: Dr Lindquist reported receiving grants from the National Institute on Drug Abuse (NIDA) during the conduct of the study and grants from the NIDA and the National Science Foundation outside the submitted work. Dr Prinstein reported receiving grants from the Winston Family Foundation during the conduct of the study. Dr Telzer reported receiving a grant from NIDA and funds from the Winston Family Foundation during the conduct of the study. No other disclosures were reported.

Funding/Support: This research was supported by National Institutes of Health grant R01DA039923 (Dr Telzer) and the Winston Family Foundation.

Role of the Funder/Sponsor: The supporters of this study had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Additional Contributions: We are extremely thankful to the families who participated in this research. We greatly appreciate the assistance of the Biomedical Research Imaging Center at the University of North Carolina at Chapel Hill, as well as Carina Fowler, BA, Susannah Ivory, BA, Amanda Benjamin, BA, Vinalia Jimenez, MS, Karoe Jones, BA, Emily Wellington, BS, Emily Bibby, BA, Rosario Villa, BA, Melissa Burroughs, BS, Kathy Do, PhD, Ethan McCormick, PhD, Paul Sharp, PhD, Lynda Lin, PhD, Nathan Jorgensen, MS, Christina Rogers, PhD, Jorien van Hoorn, PhD, Tae-Ho Lee, PhD, Natasha Duell, PhD, Jimmy Capella, MA, Maria Sobrino, BS, Zaiai Kilic, BA, Courtney Medina, BA, and Alexa Cilia, BA. At the time of their contributions to this project, all were affiliated with UNC Chapel Hill. All were compensated for their work and were either a paid research assistant, project coordinator, graduate student, or postdoctoral scholar.

REFERENCES

1. Nesi J, Choukas-Bradley S, Prinstein MJ. Transformation of adolescent peer relations in the social media context: part 1-a theoretical framework and application to dyadic peer relationships. *Clin Child Fam Psychol Rev*. 2018;21(3):267-294. doi:10.1007/s10567-018-0261-x

2. Somerville LH. Special issue on the teenage brain: sensitivity to social evaluation. *Curr Dir Psychol Sci*. 2013;22(2):121-127. doi:10.1177/0963721413476512

3. Odgers CL, Jensen MR. Annual research review: adolescent mental health in the digital age: facts, fears, and future directions. *J Child Psychol Psychiatry*. 2020;61(3):336-348. doi:10.1111/jcpp.13190

4. Prinstein MJ, Nesi J, Telzer EH. Commentary: an updated agenda for the study of digital media use and adolescent development—future directions following Odgers & Jensen (2020). *J Child Psychol Psychiatry*. 2020;61(3):349-352. doi:10.1111/jcpp.13219

5. Griffiths M. Adolescent social networking: how do social media operators facilitate habitual use? *Education and Health*. 2018;36(3):66-69.

6. Rideout V, Fox S. Digital health practices, social media use, and mental well-being among teens and young adults in the U.S. July 1, 2018. Accessed March 20, 2022. https://digitalcommons.psjhealth.org/publications/1093

7. Vogels EA, Gelles-Watnick R, Massarat N. Teens, social media and technology 2022. Pew Research Center. August 10, 2022. Accessed August 18, 2022. https://www.pewresearch.org/internet/2022/08/10/teens-social-media-and-technology-2022/

8. Blakemore SJ, Mills KL. Is adolescence a sensitive period for sociocultural processing? *Annu Rev Psychol*. 2014;65:187-207. doi:10.1146/annurev-psych-010213-115202

9. Casey BJ, Getz S, Galvan A. The adolescent brain. *Dev Rev*. 2008;28(1):62-77. doi:10.1016/j.dr.2007.08.003

10. Cohen JR, Asarnow RF, Sabb FW, et al. A unique adolescent response to reward prediction errors. *Nat Neurosci*. 2010;13(6):669-671. doi:10.1038/nn.2558

11. Ernst M, Nelson EE, Jazbec S, et al. Amygdala and nucleus accumbens in responses to receipt and omission of gains in adults and adolescents. *Neuroimage*. 2005;25(4):1279-1291. doi:10.1016/j.neuroimage.2004.12.038

12. Galván A, McGlennen KM. Enhanced striatal sensitivity to aversive reinforcement in adolescents versus adults. *J Cogn Neurosci*. 2013;25(2):284-296. doi:10.1162/jocn_a_00326

13. Van Leijenhorst L, Zanolie K, Van Meel CS, Westenberg PM, Rombouts SARB, Crone EA. What motivates the adolescent? brain regions mediating reward sensitivity across adolescence. *Cereb Cortex*. 2010;20(1):61-69. doi:10.1093/cercor/bhp078

14. Galván A. The teenage brain: sensitivity to rewards. *Curr Dir Psychol Sci*. 2013;22(2):88-93. doi:10.1177/0963721413480859

15. Steinberg L. A social neuroscience perspective on adolescent risk-taking. *Dev Rev*. 2008;28(1):78-106. doi:10.1016/j.dr.2007.08.002

16. Veissière SPL, Stendel M. Hypernatural monitoring: a social rehearsal account of smartphone addiction. *Front Psychol*. 2018;9:141. doi:10.3389/fpsyg.2018.00141

17. Rozgonjuk D, Wadsley M. A reward and incentive-sensitization perspective on compulsive use of social networking sites—wanting but not liking predicts checking frequency and problematic use behavior. *Addict Behav*. 2021;116:106808. doi:10.1016/j.addbeh.2020.106808

18. Lenhart A, Purcell K, Smith A, Zickuhr K. Social media and mobile internet use among teens and young adults: millennials. Pew Internet and American Life Project. 2010. Accessed August 18, 2022. https://eric.ed.gov/?id=ED525056

19. Cremers HR, Veer IM, Spinhoven P, Rombouts SARB, Roelofs K. Neural sensitivity to social reward and punishment anticipation in social anxiety disorder. *Front Behav Neurosci*. 2015;8:439. https://www.frontiersin.org/articles/10.3389/fnbeh.2014.00439. doi:10.3389/fnbeh.2014.00439

20. Spreckelmeyer KN, Krach S, Kohls G, et al. Anticipation of monetary and social reward differently activates mesolimbic brain structures in men and women. *Soc Cogn Affect Neurosci*. 2009;4(2):158-165. doi:10.1093/scan/nsn051

21. Chen G, Saad ZS, Britton JC, Pine DS, Cox RW. Linear mixed-effects modeling approach to FMRI group analysis. *Neuroimage*. 2013;73:176-190. doi:10.1016/j.neuroimage.2013.01.047

22. Larsen R. Missing data imputation versus full information maximum likelihood with second-level dependencies. *Struct Equ Modeling*. 2011;18(4):649-662. doi:10.1080/10705511.2011.607721

23. Ward HA, Riederer SJ, Grimm RC, et al. Prospective multiaxial motion correction for fMRI. *Magn Reson Med*. 2000;43(3):459-469. doi:10.1002/(sici)1522-2594(200003)43:3<459::aid-mrm19>3.0.co;2-1

24. Galvan A, Hare TA, Parra CE, et al. Earlier development of the accumbens relative to orbitofrontal cortex might underlie risk-taking

behavior in adolescents. J Neurosci. 2006;26(25):6885-6892. doi:10.1523/JNEUROSCI.1062-06.2006

25. Gunther Moor B, van Leijenhorst L, Rombouts SARB, Crone EA, Van der Molen MW. Do you like me? neural correlates of social evaluation and developmental trajectories. Soc Neurosci. 2010;5(5-6):461-482. doi:10.1080/17470910903526155

26. Van Leijenhorst L, Gunther Moor B, Op de Macks ZA, Rombouts SARB, Westenberg PM, Crone EA. Adolescent risky decision-making: neurocognitive development of reward and control regions. Neuroimage. 2010;51(1):345-355. doi:10.1016/j.neuroimage.2010.02.038

27. Vrtička P, Sander D, Anderson B, Badoud D, Eliez S, Debbané M. Social feedback processing from early to late adolescence: influence of sex, age, and attachment style. Brain Behav. 2014;4(5):703-720. doi:10.1002/brb3.251

28. Ordaz SJ, Foran W, Velanova K, Luna B. Longitudinal growth curves of brain function underlying inhibitory control through adolescence. J Neurosci. 2013;33(46):18109-18124. doi:10.1523/JNEUROSCI.1741-13.2013

29. Bromberg-Martin ES, Matsumoto M, Hikosaka O. Dopamine in motivational control: rewarding, aversive, and alerting. Neuron. 2010;68(5):815-834. doi:10.1016/j.neuron.2010.11.022

30. Bjork JM, Knutson B, Fong GW, Caggiano DM, Bennett SM, Hommer DW. Incentive-elicited brain activation in adolescents: similarities and differences from young adults. J Neurosci. 2004;24(8):1793-1802. doi:10.1523/JNEUROSCI.4862-03.2004

31. Galván A. Neural systems underlying reward and approach behaviors in childhood and adolescence. Curr Top Behav Neurosci. 2014;16:167-188. doi:10.1007/978-3-662-45758-0_240

32. Chambers RA, Taylor JR, Potenza MN. Developmental neurocircuitry of motivation in adolescence: a critical period of addiction vulnerability. Am J Psychiatry. 2003;160(6):1041-1052. doi:10.1176/appi.ajp.160.6.1041

33. Blum K, Sheridan PJ, Wood RC, et al. The D2 dopamine receptor gene as a determinant of reward deficiency syndrome. J R Soc Med. 1996;89(7):396-400. doi:10.1177/014107689608900711

34. Volkow ND, Wang GJ, Fowler JS, Tomasi D, Telang F. Addiction: beyond dopamine reward circuitry. Proc Natl Acad Sci U S A. 2011;108(37):15037-15042. doi:10.1073/pnas.1010654108

35. Spear LP. The adolescent brain and age-related behavioral manifestations. Neurosci Biobehav Rev. 2000;24(4):417-463. doi:10.1016/S0149-7634(00)00014-2

36. Bjork JM, Smith AR, Chen G, Hommer DW. Adolescents, adults and rewards: comparing motivational neurocircuitry recruitment using fMRI. PLoS One. 2010;5(7):e11440. doi:10.1371/journal.pone.0011440

8

Computers in Human Behavior 165 (2025) 108519

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Digitally curated beauty: The impact of slimming beauty filters on body image, weight loss desire, self-objectification, and anti-fat attitudes

Makenzie Schroeder *, Elizabeth Behm-Morawitz

University of Missouri, Columbia, MO, 65211, USA

ARTICLE INFO

Keywords:
Social comparison
Body dysmorphia
Beauty filters
Social media
Body image

ABSTRACT

The use of Augmented Reality (AR) beauty filters has been on the rise, given the advancement of technology making them more easily accessible, plentiful, and realistic. Although previous work has established beauty filters as a source of poor body image, little is known about the mechanisms for these outcomes. The current study applies social comparison theorizing to the use of beauty filters and establishes a new concept in the field: social self-comparison (i.e., the process of individuals making comparisons between their filtered image and real self-image). An online experiment of social media users ($N = 187$) was conducted to examine the effects of using a slimming beauty filter on body image and weight-related perceptions. Results indicate that comparison processes were strongest when participants used the beauty filter on their own image versus viewing someone else's filtered image, supporting the importance of examining social self-comparison processing. Overall, the results of the current study underscore the impact of beauty filter usage on body image, identifying body dysmorphia and social self-comparison as important mediators in the relationships between filter usage and body image-related outcomes, including a desire for weight loss, self-objectification, and anti-fat attitudes, among others.

Beauty filters have risen in availability and popularity, and with this rise have emerged concerns about the effects of filter use on mental health and the internalization of unrealistic beauty standards (Alsaggaf, 2021; Moreschi & Aaron, 2021). Social media filters use digital effects to modify photos and videos, and augmented reality (AR) beauty filters are specifically designed to enhance a person's appearance and attractiveness. They heighten the user's level of attractiveness by adding make-up, heightening cheekbones, or slimming and changing the proportions ... one's face. Social media can be a place of connection and fo... source of comparison and pressure to showcase an ... the self (Verrastro et al., 2020; Vogel et ... further encourages users to comp... showcase an idealized version of the ... negative outcomes on body perception...

Existing research highlights the nega... social comparison and body image (Fardo... 2020; Tiggemann & Anderberg, 2020). U... potential for social comparison with othe... idealized online version. Research suggests ... manipulating selfies is more harmful to self-... selfies (Vandenbosch et al., 2022). This sugges...

1 Beauty Filters

digital techniques to enhance one's image may have a negative influence on body perceptions. We argue that using beauty filters fosters a new form of social comparison – *social self-comparison*.

We define social self-comparison as comparing oneself to a digitally enhanced version of oneself, typically altered through a social media filter. The ... social self-comparison derives from the use of a di... ectively endorsed through social media platforms ... an image of the self to others online. As users ... in new ways through beauty filters, they may ... ance to these mediated versions of the self. In ... ers may make users prefer their filtered image ... nce. Beauty filter use has potentially severe ... influence some users to feel poorly about ... healthy cognitions and physical practices ... et al., 2022; Ren et al., 2021; Ren, ... e and technology continue to grow and ... oblem arises with what these filters ... tal expectations for what a person can ... ology Review reports that around 200 ... n Snapchat, and around 600 million ... or Instagram (Ryan-Mosley, 2021).

* Corresponding author.
E-mail address: ... (M. Schroed...

https://doi.org/10.1016/j.chb.2024.108519
Received 5 August 2024; Received in revised form 19 No...
Available online 29 November 2024

Computers in Human Behavior 165 (2025) 108519

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Digitally curated beauty: The impact of slimming beauty filters on body image, weight loss desire, self-objectification, and anti-fat attitudes

Makenzie Schroeder *, Elizabeth Behm-Morawitz

University of Missouri, Columbia, MO, 65211, USA

ARTICLE INFO

Keywords:
Social comparison
Body dysmorphia
Beauty filters
Social media
Body image

ABSTRACT

The use of Augmented Reality (AR) beauty filters has been on the rise, given the advancement of technology making them more easily accessible, plentiful, and realistic. Although previous work has established beauty filters as a source of poor body image, little is known about the mechanisms for these outcomes. The current study applies social comparison theorizing to the use of beauty filters and establishes a new concept in the field: social self-comparison (i.e., the process of individuals making comparisons between their filtered image and real self-image). An online experiment of social media users ($N = 187$) was conducted to examine the effects of using a slimming beauty filter on body image and weight-related perceptions. Results indicate that comparison processes were strongest when participants used the beauty filter on their own image versus viewing someone else's filtered image, supporting the importance of examining social self-comparison processing. Overall, the results of the current study underscore the impact of beauty filter usage on body image, identifying body dysmorphia and social self-comparison as important mediators in the relationships between filter usage and body image-related outcomes, including a desire for weight loss, self-objectification, and anti-fat attitudes, among others.

Beauty filters have risen in availability and popularity, and with this rise have emerged concerns about the effects of filter use on mental health and the internalization of unrealistic beauty standards (Alsaggaf, 2021; Montecchi & Aaron, 2024). Social media filters use digital effects to modify photos and videos, and augmented reality (AR) beauty filters are specifically designed to enhance a person's appearance and attractiveness. They heighten the user's level of attractiveness by adding make-up, heightening cheekbones, or slimming and changing the proportions of one's face. Social media can be a place of connection and fun but also a source of comparison and pressure to showcase an idealized version of the self (Verrastro et al., 2020; Vogel et al., 2014). As social media further encourages users to compare themselves with others and showcase an idealized version of the self, beauty filters may exacerbate negative outcomes on body perceptions.

Existing research highlights the negative influence of social media on social comparison and body image (Fardouly et al., 2015; Jiang & Ngien, 2020; Tiggemann & Anderberg, 2020). Using beauty filters extends the potential for social comparison with others to compare yourself to an idealized online version. Research suggests that taking and editing or manipulating selfies is more harmful to self-image than simply posting selfies (Vandenbosch et al., 2022). This suggests that the process of using digital techniques to enhance one's image may have a negative influence on body perceptions. We argue that using beauty filters fosters a new form of social comparison – social *self*-comparison.

We define social self-comparison as comparing oneself to a digitally enhanced version of oneself, typically altered through a social media filter. The "social" in social self-comparison derives from the use of a digital tool that is collectively endorsed through social media platforms and utilized to present an image of the self to others online. As users begin seeing themselves in new ways through beauty filters, they may compare their real appearance to these mediated versions of the self. In this way, using beauty filters may make users prefer their filtered image over their real-life appearance. Beauty filter use has potentially severe implications as filters may influence some users to feel poorly about their bodies and engage in unhealthy cognitions and physical practices to change their appearance (Javornik et al., 2022; Ross et al., 2021; Sun, 2021). As social media filter use and technology continue to grow and become more realistic, the problem arises with what these filters communicate to users about societal expectations for what a person can (and should) look like. MIT Technology Review reports that around 200 million daily users use AR filters on Snapchat, and around 600 million have used an AR filter on Facebook or Instagram (Ryan-Mosley, 2021).

* Corresponding author.
E-mail address: [illegible] (M. Schroeder).

https://doi.org/10.1016/j.chb.2024.108519
Received 5 August 2024; Received in revised form 19 November 2024; Accepted 27 November 2024
Available online 29 November 2024
0747-5632/

In sum, AR filters are widespread and their use rampant. Coupled with the negative impacts of their usage, it is clear that the ability of AR filters to induce social self-comparison is worth exploration, including in how this phenomenon deepens the effects of traditional social comparison.

Our research examines the effects of beauty filters on social comparison, social self-comparison, and various body-related outcomes. Participants were randomly assigned to one of three conditions: using a beauty filter (one that slims the face), watching someone else use a beauty filter, or using a filter that simply changed the color of one's photo (i.e., adding a blue film over the image). We hypothesized that using a beauty filter would result in higher levels of body dysmorphia, body ideal discrepancy, desire for weight loss, self-objectification, anti-fat attitudes, and a preference for the filtered image, mediated by comparison processes and body dysmorphia. This study utilized an experimental design to examine group differences for these outcomes based on the type of filter used, with the slimming beauty filter predicted to foster heightened negative cognitions about one's body in comparison to the other filter conditions. Social self-comparison and body dysmorphia emerged as critical factors in understanding the negative impacts of filter usage. The findings highlight the detrimental effects of AR beauty filters on body image, emphasizing the need for awareness of the potential outcomes of using such digital tools. The results of this research also advance social comparison theorizing in the age of social media.

1. Literature review

1.1. *Social comparison and social media*

Social comparison theory is a psychological framework from which to understand how the process of comparing ourselves to others helps us to better understand ourselves and our traits, such as attractiveness (Myers & Crowther, 2009). In other words, human beings make comparisons to others as a means of self-assessment. There are various ways in which social comparison theorizing has expanded over the years, including how we conceptualize and measure state-level and trait-level social comparison. State-level comparisons are those that are prompted or triggered by salient others, stimuli, and situations. For example, this may include someone reacting to an advertisement featuring a thin, stereotypically attractive model in which they then consider how their body compares (Tiggemann & McGill, 2004). Alternatively, one may just be higher in social comparison orientation, or trait-level social comparison, which is one's general propensity to make comparisons to others (Gibbons & Buunk, 1999).

When using Instagram and other social media platforms, users may engage in upward comparison by comparing themselves to someone perceived to be better off in some way (in this case, based on the appearance of their body), leading to a drive for thinness, body dissatisfaction, body surveillance, and other appearance-related concerns (Serkis et al., 2020). These comparisons may also negatively impact health, including prompting disordered eating (e.g., Saunders & Eaton, 2018). Traditionally, upward comparisons on social media have been made with others on those platforms, such as models (e.g., Tiggemann & McGill, 2004), celebrities, and friends (e.g., Ho et al., 2016). Social media posts guised in the context of "fitness" and "health" content have also been demonstrated to increase social comparison and negative body image for women, as cumulative exposure to these posts exposes users to a slim, idealized version of a fit body (Lewallen & Behm-Morawitz, 2016). However, with the advancement of social media has also come the advancement of other related technologies available on these platforms, such as augmented reality (AR) beauty filters.

1.2. *Filter usage and body image*

Through filters, users can rapidly and efficiently picture themselves in new, digitally altered ways, given the advancement of AR technologies – with a different head shape, eye color, and even facial structure, size, and features. It is no secret in body image research that exposure to manipulated/retouched photos can hurt the body image of viewers – this becomes even stronger for users high in social comparison tendencies (Kleemans et al., 2018). One review of the literature found that most studies in this domain show a positive relationship between photo editing and body concerns as well as related outcomes, such as self-objectification, body shame, and overvaluation of weight (McGovern et al., 2022).

Even more importantly, filters and photo editing allow users to see themselves in ways they may even find preferential to their current, real selves. This ability comes with serious implications. For example, high-frequency users of filters on Instagram are more likely to engage in social appearance comparison and internalize general attractiveness ideals (Mancin et al., 2024). This draw to filters may stem from a desire for idealized self-presentation, as filters allow users to "fix" their insecurities, like blurring a pimple or whitening teeth (Javornik et al., 2022). As people begin to see themselves more desirably, they may begin to wish for those changes permanently. Therefore, it is no surprise that the use of filters and photo editing has been linked to greater acceptance of cosmetic surgery as well as greater intentions to have cosmetic surgery (Beos et al., 2021; Sun, 2021).

1.3. *Expanding social comparison theory*

1.3.1. *Social self-comparison*

We extend social comparison theorizing to examine the comparative processes users may engage in when viewing their images altered by AR beauty filters. Specifically, users may compare themselves to a "better" version of themselves – their filtered image. As explained earlier, we conceptualize this as social self-comparison – comparing oneself to a digitally enhanced version, typically altered through an AR social media filter for images or videos. Research shows that typical upward comparisons to models, peers, and other people can result in negative body image (e.g., Serkis et al., 2020). Therefore, one may expect that making comparisons to a closer object (in this case, an idealized version of oneself) will also result in a comparative cognitive process that produces negative body perceptions. Social self-comparison provides researchers a lens through which they can study the influence of filters, selfies, and social media documentation of the self on perceptions of the self. The process is inherently social in nature, given the communal use of filters, the public or semi-public nature of social media, and the sharing of social media content.

In short, social self-comparison is a new theorizing about comparison, highlighting that the object of comparison can be a digitally altered version of the self. This may be especially important in the context of how users feel about their bodies and view them, as each type of comparison may serve as a mechanism for other negative outcomes. This may include exaggerating the gap between users' current weight and how they wish to look (body ideal discrepancy) as well as their desire to lose weight. Specifically, in the case of this study, state-level social comparison will increase after exposure to someone else using a beauty filter, leading to these outcomes. Similarly, the same process will happen as other participants use a beauty filter. However, this indirect effect will be much weaker in the case of using a filter that does not change a person's appearance to be more socially desirable.

How much participants view themselves as an object and value their own sexual appeal, known as self-objectification (Fredrickson & Roberts, 1997), may also be impacted by beauty filters. Self-objectification is about seeing oneself from a third-person perspective, placing looks above ability and feelings in importance. In short, what matters to our self-concept is the perception of how others view their bodies in terms of sexual appeal and attractiveness. Therefore, a consequence of self-objectification is body surveillance, or the constant thinking about and monitoring of one's appearance. Additionally, self-objectification can result in the experience of body shame, appearance anxiety, a "diminished awareness of internal bodily states" like being able to

identify hunger and emotions, as well as "reduced concentration" on "mental and physical tasks" (Calogero, 2012, p. 575). In instances where self-objectification is heightened, like during a college student organization recruitment event or while trying on clothes, people may also engage in disordered eating, experience increased sexual dysfunction, body shame, and even worse performance in school (Rolnik et al., 2010; see Daniels et al., 2020 for a review; see Moradi & Huang, 2008 for a review). Given this, self-objectification can operate as a trait, but also may occur in moments where the body and its appearance is centered (Calogero, 2012). In this way, self-objectification can be triggered, in this case by making comparisons to filtered social media images.

Additionally, the use of beauty filters and comparison processes may lead to negative views about fatness, or anti-fat attitudes. Anti-fat attitudes are rooted in fatphobia, which is "a pathological fear of fatness often manifested as negative attitude and stereotypes about fat people" (Robinson et al., 1993, p. 468). Fatphobia is pervasive in society, both in one's view of the self and others. In one study, women were asked to imagine gaining 100 pounds said they would consider moving away or even suicide (Fahs & Swank, 2017). In another study, nearly half of respondents reported that they'd be willing to give up one year of their life rather than be obese. Some respondents also indicated they'd be willing to give up 10 or more years of life (15%), lose the ability to have children (25%), or get divorced (30%) (Schwartz et al., 2006). This provides a clear indication of just how deeply rooted fatphobia is within US society, as it is avoided at devastatingly large costs.

In using filters, individuals can see themselves in ways that uphold societal standards for beauty, including thinness. Users may seek or desire permanent cosmetic changes as they begin to picture themselves in these new ways (Beos et al., 2021; Sun, 2021). In turn, they may begin to develop a dislike of their current appearance. In the case of beauty filters, the filtered image and one's real appearance may result in a greater dislike of their excess weight. Therefore, they may experience greater anti-fat attitudes and a preference for the filtered image. Given the aforementioned information, we propose the following.

Hypothesis 1. Condition will predict social comparison, such that those using a beauty filter will experience the greatest level of comparison followed by those watching someone else use a beauty filter and the control condition.

Hypothesis 2. State comparison will mediate the relationship between condition and a) body ideal discrepancy, b) desire for weight loss, c) self-objectification, d) anti-fat attitudes, and e) preference for filtered image.

See Fig. 1 for a pictorial representation of Hypothesis 2 predicted mediation relationships. Greater levels of state comparison will foster greater levels of the outcome variables.

1.3.2. *Body dysmorphia*

Body dysmorphia, or one's perception of their body as being flawed beyond what an objective outsider may see, is well-established as being a predictor of disordered eating and exercise behavior and self-criticism (Alsaidan et al., 2020; Matos et al., 2023). Symptoms of body dysmorphia include increased body consciousness, shame, guilt, and perceived body surveillance (Fanaghi et al., 2019; du Roches et al., 2023). For those experiencing body dysmorphia, they may feel embarrassed about their bodies and, therefore, avoid social situations and engage in disordered eating (Rizwan et al., 2022). Body dysmorphia may have additional serious implications for physical and mental health, as body dysmorphic disorder has been linked to cosmetic surgeries and even self-mutilation to be approved for otherwise healthy limb removal in extreme cases (Chan et al., 2011). Social media has been linked to body dysmorphia in that social media use may lead to an increase in dysmorphia and related outcomes, such as poor self-esteem and anxiety (Rizwan et al., 2022; Raj et al., 2022). This includes the use of filters and other photo editing online.

Beauty filter usage may increase body dysmorphic cognitions, as seeing an idealized version of others and themselves may make individuals feel negatively about their real-life looks. Given the connection between body dysmorphia and a desire to change one's appearance (Rizwan et al., 2022; Chan et al., 2011), filter users may feel unsatisfied with their current selves after filter use, and wish to make an appearance change. In short, using a beauty filter or watching someone else use a beauty filter may trigger someone to feel unreasonably poorly or anxious about their current body (i.e., body dysmorphic thoughts). Although short-term social media filter exposure may not elicit change in body dysmorphic behaviors, it may trigger body dysmorphic thoughts. Over time, body dysmorphic cognitions may result in disordered behavior. As such, we examine the relationship between social media filter exposure on short-term body dysmorphic cognitions. Due to heightened surveillance of the body and feeling poorly about their current appearance, individuals' desire for weight loss, feeling disgust toward fatness, and the like, are predicted to increase. These effects are expected to emerge most strongly for condition 1 where participants are using the slimming beauty filter on an image of themselves.

Hypothesis 3. Condition will predict body dysmorphic cognitions, such that those using a beauty filter will experience the greatest level of body dysmorphia followed by those watching someone else use a beauty filter and the control condition.

Hypothesis 4. Body dysmorphic cognitions will mediate the relationship between condition and a) body ideal discrepancy, b) desire for weight loss, c) self-objectification, d) anti-fat attitudes, and e) preference for filtered image.

See Fig. 2 for a pictorial representation of Hypothesis 4 predicted mediation relationships. Greater levels of body dysmorphic cognitions will foster greater levels of the outcome variables.

2. Method

2.1. *Procedure and participants*

An online experiment examined the effects of a beauty filter on body image-related outcomes. Our experiment consisted of three conditions (see Table 1) via random assignment: condition one tasked participants with using a filter that slimmed their face ($n = 63$); condition two consisted of participants watching someone else use a beauty filter, but not using it themselves ($n = 79$); and the third condition required participants to use a neutral filter unrelated to body size/shape. This filter changed the color of their screen/image to blue ($n = 45$). The last condition served as a control condition.

Participants were recruited via ResearchMatch, a non-profit recruitment service consisting of volunteer participants for health-

Fig. 1. Pictorial representation of Hypothesis 2.

Fig. 2. Pictorial representation of Hypothesis 4.

Computers in Human Behavior 165 (2025) 108519

Table 1
Descriptions of experimental conditions.

Condition	Description
1	Participants used a slim beauty filter on image of self
2	Participants watched a video of a person using a slim beauty filter
3	Participants used a color filter on image of self

related studies in the United States. The service is funded by the National Institutes of Health (NIH). We limited our recruitment to English-speaking adults, primarily under the age of 50, to best reflect the demographic most familiarized and connected to the use of social media filters. Upon recruitment and consent to participate in the study, participants were asked to answer questions regarding their social media consumption and their perceived body size. Afterward, they were given detailed instructions to download the free app, Snow, where they were then tasked with using a specific filter, depending on their condition. This app was chosen due to its ease of accessibility given its cost-free access, as well as its widespread use, as the company reports hosting over 200 million users. We believe that their offerings were representative of filters across social media platforms. Additionally, the app has a large selection of filters, allowing for both a slimming filter and neutral filter – both of which were needed for the study.

Participants in Condition 1 were instructed to use the same beauty filter; participants in Condition 2 were shown a video of someone using the same beauty filter, but did not use it themselves; and in the third condition, participants were instructed to use the color-changing filter. In the condition directions, participants were given screenshots of each step, including the filter to select. The full instructions for each condition are available in the Supplemental Appendix. After condition exposure, participants completed all other measures used in the study's analysis.

The study's true purpose was disguised by the use of general language in the consent form about the study's aims, as well as including distractor items such as health efficacy measures in the questionnaire. We also employed neutral language in our measures to avoid prompting a social desirability bias from participants.

As a data quality check, all participants who indicated that they used other filters (beyond the one they were instructed to use, $n = 4$) during the study were removed from the analysis, as well as those who did not complete the survey, leaving a final sample of $N = 187$. The majority of the participants were white ($n = 153$), followed by Black ($n = 12$), with the remaining participants identifying as another race or as multiple races. The sample consisted of a majority women ($n = 156$), followed by men ($n = 16$), trans women ($n = 13$), and trans men ($n = 2$). The age of participants ranged from 19 to 66 years old ($M = 36.28$, median = 35, $SD = 10.12$).

2.2. *Measures*

2.2.1. *Body ideal discrepancy*

Using images adapted from Stunkard et al. (1983), to measure body ideal discrepancy we first showed them illustrations of nine body size types, ranging from the (1) smallest body to the (9) largest. We asked participants to identify their current perceived body size before using the filter, and then the body size they would prefer to have after using the filter. From these responses, we subtracted the score of one's ideal body from their current perceived body size score. The resulting value is what we conceptualized as body ideal discrepancy, as we were interested in determining how different their current body is from the body they wish they had ($M = 1.4$, $SD = 1.37$).

2.2.2. *Body dysmorphic cognitions*

To capture participants' body dysmorphia, we utilized six items from the cognitions subscale of the Body Dysmorphia Scale (BDD-SS) (Wilhelm et al., 2016). The cognitions subscale measures the internal dialogue one may have regarding concerns over their appearance. These cognitions, more so than symptomatic behaviors of body dysmorphia (e.g., exercising excessively, avoiding mirrors), may become salient after using a slimming filter. Items included "If I could look just the way I wish, I would be much happier" and "If my appearance is defective, I will end up alone and isolated." Responses ranged from (1) *strongly disagree* to (7) *strongly agree* ($M = 3.4$, $SD = 1.27$, $\alpha = 0.82$). Higher response scores represent a higher occurrence of body dysmorphia in participants.

2.2.3. *Desire for weight loss*

In order to evaluate participants' desire for weight loss, we presented three items: "I would like to lose weight," "I am happy with my current weight" (reverse-coded), and "It is my goal to lose weight." Responses ranged from (1) strongly disagree to (7) strongly agree, with a higher score indicating a greater desire to lose weight ($M = 4.98$, $SD = 1.72$, $\alpha = 0.85$).

2.2.4. *Self-objectification*

To measure self-objectification, we utilized the Self-Objectification Questionnaire from Noll and Fredrickson (1998). Although initially identified as a trait measure, it has been used effectively in numerous experimental and panel studies (e.g., Aubrey, 2006; Aubrey et al., 2009; Petonll & Behm-Morawitz, 2015) to demonstrate state-level change in self-objectification, due to priming of a stimulus (e.g., slimming filter image exposure). This measure asks participants, "How important are these attributes to your self-concept? From most important (1) to least important (10), please rank the following attributes." These attributes include five appearance/desire-based attributes, like weight and sex appeal, as well as five competence-based attributes, like physical coordination and energy level. To calculate the total score on participants' objectification tendencies, we subtracted the sum of the appearance/desire measures from the competence measures. Therefore, within the composite measure, a lower score indicated greater self-objectification ($M = 1.89$, $SD = 2.66$).

2.2.5. *Anti-fat attitudes*

To assess anti-fat/fatphobic attitudes, which is the extent to which people hold negative attitudes toward overweight individuals and personal weight gain, we utilized Crandall's (1994)Anti-fat Attitudes Scale. This included a total of 13 items across three subscales (dislike, fear of fat, and willpower), such as "I really don't like fat people much" and "Fat people make me feel somewhat uncomfortable." Responses ranged from (1) strongly disagree to (7) strongly agree), with a higher score indicating greater fatphobic attitudes ($M = 2.97$, $SD = 1$, $\alpha = 0.85$).

2.2.6. *Social media use*

As a control variable, we were interested in measuring participants' social media usage, as those higher in social media use may be more susceptible to the effects of social comparison, including with filter usage as they experience greater exposure. Therefore, to capture this, we asked participants, "Approximately, how many hours per day do you spend interacting with social media (e.g., TikTok, Instagram, Twitter, Facebook, etc.)?" They were then able to select zero hours to over 10 h from a drop-down menu ($M = 2.6$, $SD = 2.32$).

2.2.7. *State comparison*

State comparison measured both social self-comparison and social comparison, depending on the condition. Social self-comparison was captured using an adapted version of Tiggemann and McGill's (2004) state appearance comparison scale. Three items were constructed to match the stimuli of each condition. For example, the social comparison questions in condition one ($M = 4.85$, $SD = 1.89$, $\alpha = 0.85$) and three ($M = 3.94$, $SD = 1.78$, $\alpha = 0.85$) included: To what extent did you think about your unfiltered appearance when seeing your filtered appearance? To what extent did you compare your overall unfiltered appearance with your filtered appearance? To what extent did you compare unfiltered

specific body parts (like your face) with your filtered appearance? Questions in the second condition ($M = 2.68$, $SD = 1.96$, $\alpha = 0.85$) included traditional social comparison items: To what extent did you think about your own appearance when watching the person using a filter? To what extent did you compare your overall appearance with the person using a filter? To what extent did you compare specific body parts (like your face) with that of the person using a filter?

3. Results

3.1. Social comparison and social self-comparison

First, H1 was tested using analysis of variance to decompose the direct effects of condition on state comparison (i.e., social comparison and social self-comparison) in order to examine significant group differences. Social media use was entered as a covariate in the model. Results revealed that the condition (see Table 1) significantly predicted the level of state comparison, $F(1, 185) = 23.20$, $\eta_p^2 = 0.20$, $p < .001$ (see Table 2). To further understand the extent of difference between conditions, Tukey HSD post hoc tests revealed significant differences between conditions 1 and 2 (*mean difference* = 2.16, $p < .001$), conditions 1 and 3 (*mean difference* = 0.91, $p < .05$), and conditions 2 and 3 (*mean difference* = −1.26, $p < .001$). In other words, state comparison was greatest for those in condition 1, followed by condition 3 and then condition 2. State comparison in condition 1 (i.e., using the slim beauty filter) was measured using social self-comparison, the new construct proposed in this research. Our finding suggests viewing a filtered image of the self resulted in greater comparison processes than viewing a filtered image of another person.

Next, hypothesis 2 examined state-level comparison as a mediator of the relationship between condition and the outcome variables. The indirect effects models were tested using Hayes' PROCESS version 4.2 model 4. In the models, Condition 1 was used as the referent condition, and two contrasts were tested: Condition 2 vs. Condition 1 (X1) and Condition 3 vs. Condition 1 (X2). Significant indirect effects models are indicated by Lower Limit Confidence Intervals (LLCI) and Upper Limit Confidence Intervals (ULCI) that do not contain zero. Social media use was entered as a covariate. State comparison mediated the relationships between condition and the following outcomes: self-objectification, $F(4, 182) = 4.03$, $R^2 = 0.08$, $p < .001$; desire to lose weight, $F(4, 182) = 2.97$, $R^2 = 0.06$, $p < .05$; and anti-fat attitudes, $F(4, 182) = 5.68$, $R^2 = 0.11$, $p < .05$. Please refer to Table 3 for full indirect effects model results examining condition differences.

Conditions 2 (i.e., viewing slimming filter on other person) and Condition 3 (i.e., using color filter on self) both significantly differed from Condition 1 (i.e., using slimming filter on self), such that Condition 1 produced greater comparison levels than the other conditions, which in turn significantly predicted more negative outcomes. In short, social self-comparison—the novel form of social comparison developed in our research—fostered more negative attitudes toward fatness, higher levels of self-objectification, and desire to be thinner.

Table 2
Levels of body dysmorphic cognitions and state comparison predicted by condition.

Variable	Condition 1 Mean (SD)	Condition 2 Mean (SD)	Condition 3 Mean (SD)
Body dysmorphic cognition	3.63 (1.37)[a]	3.38 (1.26)	3.09 (1.08)[a]
State comparison	4.85 (1.84)[bc]	2.68 (1.96)[bd]	3.94 (1.78)[cd]

[a]Conditions 1 and 3 were significantly different for body dysmorphic cognitions.
[bc]Condition 1 was significantly different from both Conditions 2 and 3 for state comparison.
[bd]Condition 2 was significantly different from both Conditions 1 and 3 for state comparison.
[cd]Condition 3 was significantly different from both Conditions 1 and 2 for state comparison.

Table 3
Indirect effects of condition on dependent variables through state comparison.

Dependent Variable	Contrast	b	SE	LLCI	ULCI
Body ideal discrepancy	X1	0.148	0.12	−0.083	0.392
	X2	0.065	0.06	−0.037	0.198
Self-objectification	X1*	0.572	0.26	0.125	1.122
	X2*	0.262	0.15	0.025	0.590
Preference for filtered image	X1	−0.031	0.11	−0.278	0.179
	X2	−0.022	0.05	−0.141	0.082
Desire to lose weight	X1*	−0.406	0.16	−0.756	−0.118
	X2*	−0.179	0.10	−0.414	−0.024
Anti-fat attitudes	X1*	−0.271	0.10	−0.482	−0.101
	X2*	−0.13	0.06	−0.265	−0.025

Note. The indirect effects model results for X1 indicate the difference between Condition 1 and Condition, and the results for X2 indicate the difference between Condition 1 and Condition 3. Condition 1 (i.e., using the slimming beauty filter) was indicated as the reference group.
* Denotes significant total indirect effects for the outcome variable.

However, state comparison did not mediate the relationships between filter usage and body ideal discrepancy, nor preference for the filtered image; meaning engaging in state comparison to the filtered image did not create a larger gap between one's current body size and ideal body size, nor did it make people wish they looked like their filtered image. However, there was a direct effect of condition on preference for the filtered image, $t = 4.88$, $p < .001$ (LLCI = 0.3686, ULCI = 0.8683). Analysis of variance was utilized to examine differences by condition. Condition significantly predicted preference for the filtered image, $F(2, 186) = 19.62$, $p < .001$, $R^2 = 0.17$. Tukey HSD post hoc tests revealed significant differences between conditions 1 and 2 (*mean difference* = −1.20, $p < .001$), conditions 1 and 3 (*mean difference* = −1.22, $p < .001$). There were no significant differences between conditions 2 and 3 (*mean difference* = −0.02, $p = .99$). Taken together, these results suggest that participants who used the slimming filter preferred that image to their real self to a greater degree than those who used the neutral color-changing filter and those who watched someone else use the slimming filter.

3.2. Body dysmorphic cognitions

Hypothesis 3 examined the direct effect of condition on body dysmorphic cognitions. Results revealed that condition significantly predicted level of body dysmorphic cognitions ($F(2, 184) = 14.26$, $R^2 = 0.13$, $p < .001$), $t = -2.63$, $p < .01$ (LLCI = −0.5315, ULCI = −0.0758). Mean scores revealed that participants who used the slim beauty filter had the highest level of body dysmorphic cognitions, followed by participants who watched a video of another person using a slim beauty filter, and finally, those who used a color filter (see Table 2). However, not all group differences were significant. Results revealed that conditions 1 and 2 as well as conditions 2 and 3 were not significantly different from one another. However, conditions 1 and 3 were significantly different from one another, such that body dysmorphic cognition was higher after participants used a slim beauty filter in comparison to using a color filter, $t(106) = 2.21$, $p < .05$).

The indirect effects models for hypothesis 4 were tested using Hayes' PROCESS version 4.2 model 4. In the models, Condition 1 was again used as the referent condition, and two contrasts were tested: Condition 2 vs. Condition 1 (X1) and Condition 3 vs. Condition 1 (X2). Significant indirect effects models are indicated by LLCI and ULCI that do not contain zero. Social media use was again entered as a covariate in the model. The relationships between condition and all outcome variables were significantly mediated by body dysmorphic cognitions, but only when looking at differences between Conditions 1 and 3 (X2). Please see Table 4 for full results. Using the slimming filter fostered higher body dysmorphic cognitions, in comparison to using the color filter, which in turn resulted in stronger body ideal discrepancy, desire for weight loss, self-objectification, preference for the filtered image, and anti-fat

Table 4
Indirect effects of condition on dependent variables through body dysmorphic cognition.

Dependent Variable	Contrast	*b*	SE	LLCI	ULCI
Body ideal discrepancy	X1	0.071	0.07	−0.063	0.223
	X2*	0.192	0.09	0.047	0.388
Self-objectification	X1	0.198	0.19	−0.145	0.594
	X2*	0.526	0.22	0.144	0.983
Preference for filtered image	X1	−0.047	0.05	−0.162	0.032
	X2*	−0.126	0.07	−0.281	−0.015
Desire to lose weight	X1	−0.128	0.12	−0.381	0.109
	X2*	−0.344	0.14	−0.634	−0.100

Note. The indirect effects model results for X1 indicate the difference between Condition 1 and Condition, and the results for X2 indicate the difference between Condition 1 and Condition 3. Condition 1 (i.e., using the slimming beauty filter) was entered as the reference group.
* Denotes significant total indirect effect for the outcome variable.

attitudes.

There were no significant differences between Conditions 1 and 2 when examining the indirect effects of condition on our outcome variables through body dysmorphic cognitions. Taken together, the results of Hypothesis 4 indicate that that using a slimming filter on an image of oneself was associated with more negative outcomes in comparison to using a neutral filter.

4. Discussion

Social self-comparison is a novel concept to formalize the everyday experience afforded by digital advancements in AR technology – making comparisons to your mediated, filtered self possible. With the prominence of social media in our daily interactions, filter usage has a social element in two ways. First, filters allow users to engage in conforming to societal expectations around beauty. People can use beauty filters that make them appear thinner, have smoother skin and longer lashes, and other standards for the idealized appearance (Eshiet, 2020). Second, it is a social experience due to the nature of social media, resulting in a shared experience of using filters that other people are using and posting filtered images and videos for other users to see. In this way, users are also thinking about the imagined audience – in turn, a cycle is created of posting filtered images and expecting filtered images of others (Lawrence & Cambre, 2020). These expectations may feed into the thin idealized beauty standards in the US among these social media users.

It is no secret then that filters create expectations around beauty and how to present oneself online (Eshiet, 2020). As our study shows, they also have the ability to deepen the stigma around fatness. While much work has been done on social comparison on social media (e.g., Verduyn et al., 2020), little is conceptualized about the related phenomenon of comparing oneself to a digitally altered version of the self. Therefore, our current study pushes forward thinking about how to expand theorizing of social comparison to include social self-comparison to examine how digitally altered images of the self impact self perceptions.

As our results highlight, using a filter was more potent in participants' subsequent state comparisons. In other words, participants using a filter were more likely to make comparisons (in this case, to their *filtered* image) than those watching someone else use a filter. This provides evidence of our proposed extension to social comparison theory. It is clear that beauty filters have created deeper opportunities for state comparisons, highlighting the importance of considering social self-comparison in body image research. This is especially important as we consider the outcomes of comparison in the digital environment.

The current study found that, when experiencing state-level comparison as a result of filter use, users felt worse about their real-life image. Specifically, the process of social self-comparison—or comparing one's real self to an augmented version of the self via a social media filter—fostered more negative effects. This manifested specifically in their desire to lose weight after seeing themselves in a slimmer way. They also were more likely to view their body as an object, evaluating themselves and their worth based on their appearance. Additionally, they experienced a greater disdain for fatness. Therefore, filter usage resulting in state comparison (via social self-comparison) has negative implications for how users subsequently feel about themselves and their bodies. Although these effects are not large, they are statistically significant and account for some variance in explaining why individuals may feel poorly about themselves after using social media filters.

The dependent variables measured in this study also are linked with other negative health outcomes, thus suggesting that future research may further explore other impacts of using social media beauty filters. For example, engaging in self-objectification may result in increased sexual dysfunction and disordered eating, among other negative outcomes (Rolnik et al., 2010; see Daniels et al., 2020 for a review; see Moradi & Huang, 2008 for a review). Additionally, with filter users experiencing a greater disdain toward fatness as a result of comparisons, this may lead to the deepening of weight stigma and dislike toward overweight individuals. Therefore, filter usage may have additional alarming consequences. Future research may examine such outcomes.

Beyond state comparison, body dysmorphic cognition was an important mediator in the relationship between filter consumption and body image-related outcomes. Although traditionally thought of in a clinical sense via body dysmorphic disorder (e.g., Phillips, 1998), body dysmorphia may operate similarly to self-objectification in that it is indeed a trait, but can be made salient in certain situations. Additionally, we intentionally only measured this using the cognitions subscale of the Body Dysmorphia Scale (BDD-SS). This omitted behavioral outcomes of body dysmorphia, as we were instead interested in participants' internal thoughts that they may be ruminating on as a result of the filter usage. These cognitions may be more widely experienced outside of a clinical sense given their closeness to other measures such as body dissatisfaction.

Our results found that body dysmorphic feelings were indeed primed by filter usage, such that filter usage led to participants experiencing a saliency effect, or heightened body dysmorphia. By experiencing body dysmorphic feelings as a result of the filter, participants were further impacted to experience great body ideal discrepancy, meaning they had a larger gap between their current body size and ideal body size. Unsurprisingly then, they also preferred their filtered image over their real-life appearance and had a greater desire to lose weight. They also were more likely to objectify themselves. Additionally, they experienced greater anti-fat or fatphobic attitudes.

Regarding differences in the mediation effects for each condition, using filters on the self resulted in higher levels of comparison, as compared to watching someone else use a filter. This is as expected, given that the mediated self is a closer object for comparison. Additionally, those using the beauty filter had higher levels of body dysmorphic cognitions compared to those using the color filter, although there were no significant differences from the body dysmorphic thoughts experienced by those watching someone else use a filter. This is unsurprising given that watching someone alter their appearance in a desired way, or altering your own appearance, is more likely to result in experiencing body-related concerns (Eshiet, 2020; Javornik et al., 2022; Barni et al., 2021; Sun, 2021). Both using and watching others use the slimming beauty filter results in more pronounced anti-fat attitudes, suggesting that filters not only promote comparison processes but also heighten negative beliefs about overweight or fat individuals. Our research findings suggest that slimming filters may have a negative effect on people's body-related perceptions as well as cultivate a narrow view of beauty that does not include larger-sized individuals.

4.1. Theoretical implications

The current study advances social comparison theorizing to include a

novel concept: social self-comparison. As the study results confirm, social self-comparison is a powerful mechanism that may be triggered by beauty filters, leading to a slew of negative outcomes for users. In this way, social media and the advancement of AR have begun to allow users easily accessible and often realistic ways to view themselves, including in desirable ways. Through using beauty filters, users are able to make comparisons between their real-life appearance and their mediated image. As results indicated, social comparison was greatest amongst those in the social self-comparison condition, followed by the condition in which someone else was using a filter. Therefore, the current study continues to find support for social comparison and its effects, but further expands our understanding of comparison processes to showcase the strength of social self-comparison.

Previous social comparison literature has focused on making comparisons to others, leading to either positive or negative self-evaluations (e.g., Buunk & Gibbons, 2007). In the current study design, not only are people engaging in self-comparisons through filters, but often walking away likely to have experienced negative self-evaluations. For example, the participants tasked with using a slimming beauty filter were able to see thinner, filtered images of themselves, leading them to want to lose weight and even feel a greater dislike toward fatness. They also engaged in more self-objectification, highlighting the pervasiveness of filter usage in appearance anxiety and self-concept (e.g. Calogero, 2012).

Recent research shows that beauty filters decrease body satisfaction, and the current study expands the ways in which that dissatisfaction may manifest (Dijkslag et al., 2024). As Dijkslag and colleagues (2024) note, this finding may be due to social comparison processes, although they did not directly evaluate this. However, the current study is able to formally solidify social comparison processing in the effects of beauty filters, as well as expand the field of social comparison theorizing. Although the study by Dijkslag and colleagues (2024) did not see support for the role of self-identification in the usage of beauty filters and their associated outcomes, the current study supports the idea that social self-comparison may be more powerful than traditional social comparison, showcasing the need for further exploration in this vein.

4.2. Practical implications

Filters are present in our everyday lives, and using them has become ingrained into the experience of digital life in society. Although some filters may seem silly or harmless, others may lead to detrimental consequences. Some of the practical implications of this study include the expectations set by slimming filters for users, as they picture themselves how they could look having lost weight. From this filter usage, users are led to compare themselves to their filtered image as well as experience heightened body dysmorphic feelings. In turn, filters lead users to feel poorly about themselves and wish to change their appearance through weight loss for appearance reasons, rather than health reasons. There may be additional practical (and theoretical) implications regarding stigma, as slimming filter usage causes a deeper dislike of fatness. In turn, slimming and other beauty filters also elicit ethical considerations, as users may feel poorly about themselves and others as a result of their usage. Social media developers may learn from this research that slimming filters are a detriment to users given their psychological harm. We recommend to practitioners that they practice ethical design and have AI/IT designers and User Experience (UX) researchers consider the empirical research that points to the negative effects of beauty filters. In turn, developers may consider eliminating these types of filters, as well as social media platforms, instead offering users more body-neutral options.

The results also expand our understanding of beauty filters in making salient experiences of body dysmorphic thoughts. As mean scores indicated, body dysmorphic cognition was greatest amongst those in the social self-comparison condition, followed by the condition in which someone else was using a filter. Although only the social self-comparison condition and control were statistically significantly different from one another, results show that whether by using a filter or watching someone else use one, body dysmorphic thoughts may become salient, leading to other negative body-related evaluations. In other words, seeing oneself "beautified" leads to the self-perception of one's appearance being flawed. The same thing occurs in seeing someone else using a filter. This further builds on research highlighting social media as a source for making body dysmorphia salient as well as causing related outcomes (Rizwan et al., 2022; Raj et al., 2022).

4.3. Limitations and future directions

Limitations of this research include the generalization of our findings to the multitude of real-world scenarios that may be encompassed in using beauty filters in the social media landscape. We tested one slimming filter that has a relatively subtle effect, whereas other filters popular on TikTok and Instagram may produce other effects and more profound effects on digital images and videos. It is not unreasonable to assume that beauty filters, collectively, may have some similar effects on social media users. However, differential effects may also emerge for various digital manipulations afforded by social media filters. We also rely on self-report data, which may not fully capture the extent of filter usage. Additionally, we employed an experimental design that lacks ecological validity; in this way, participants did *not* get to choose what filter they used. In turn, the study's external validity may not harness the impact of real-world implications. For example, not all participants may want to use the study's filters in the real world, lessening the robustness of conclusions drawn by the results.

Additionally, the average age of our participants was 36 which does not represent Generation Z social media users. Rather, the majority of our participants were Millennials who are engaging with filters alongside younger generations but potentially from different life experiences and with different effects. Further, the majority of our sample consisted of women. Future research should look further into how slimming filters, or filters that show heightened muscularity, impact men. Additional research may also seek to replicate this study with additional social media audiences to determine how effects may be similar or different based on generation and culture. Finally, our research provides a foundation for the study of social self-comparison. Future research should further unpack specific instances in which social self-comparison may be stronger.

5. Conclusion

Taken together, the results of our research support the notion that beauty filters may have a negative effect on body image, fostering internalization of unrealistic beauty standards. This is consistent with prior studies (Almegbel, 2021; Maroschi & Aaron, 2024) and extends knowledge of the processes at play in determining the effects of filter usage on cognitive outcomes. Specifically, using a slimming beauty filter resulted in greater comparison processes than watching another person use a beauty filter or using a simple color (blue) filter on one's image. Additionally, using a beauty filter resulted in greater comparison processes than using a color filter. The relationships between beauty filter use and body-related perceptions may be complex. Comparison processes mediated the relationship between condition and desire to lose weight, self-objectification, and anti-fat attitudes. However, comparison processes did not predict body ideal discrepancy or preference for the filtered image. In other words, a person's level of social/self-comparison after condition exposure did not predict the gap between their current and preferred body size, or their predilection for the filtered image. This may be because comparison is not the correct mechanism to examine the relationship, but instead, these outcomes may be more directly related to how a person feels about their appearance, such as body dysmorphia.

Our study highlights the dangerous relationship between filter use and body dysmorphia, as this relationship serves as a mechanism for negative body-related outcomes and a pressure to be more like one's

filtered image. Body dysmorphia mediated the relationships between condition and all outcome variables, suggesting that using slimming beauty filters may negatively impact people's body-related perceptions in terms of body dysmorphia and then, in turn, result in greater body dissatisfaction and negative body-related beliefs.

CRediT authorship contribution statement

Makenzie Schroeder: Writing – original draft, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Elizabeth Behm-Morawitz:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Formal analysis, Data curation.

Declaration of competing interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chb.2024.108519.

Data availability

The authors do not have permission to share data.

References

Alnaggar, R. M. (2021). The impact of Snapchat beautifying filters on beauty standards and self-image: A self-discrepancy approach. ISSN: 2188-1111 – The European Conference on Arts & Humanities 2021: Official Conference Proceedings https://doi.org/10.22492/issn.2188-1111.2021.4.

Alsaidan, M. S., Alotaibi, N. S., Alzhrani, S. H., Alshammari, M. M., Alqahtani, F. T., & Mohajer, K. A. (2020). The prevalence and determinants of body dysmorphic disorder among young social media users: A cross-sectional study. Dermatology Reports, 12(3), 8774. https://doi.org/10.4081/dr.2020.8774

Aubrey, J. S. (2006). Effects of sexually objectifying media on self-objectification and body surveillance in undergraduates: Results of a 2-year panel study. Journal of Communication, 56(2), 366–386. https://doi.org/10.1111/j.1460-2466.2006.00024.x

Aubrey, J. S., Henson, J. R., Hopper, K. M., & Smith, S. E. (2009). A picture is worth twenty words (about the self): Testing the priming influence of visual sexual objectification on women's self-objectification. Communication Research Reports, 26(4), 271–284. https://doi.org/10.1080/08824090903293551

Beos, N., Kemps, E., & Prichard, I. (2021). Photo manipulation as a predictor of facial dissatisfaction and cosmetic procedure attitudes. Body Image, 39, 194–201.

Buunk, A. P., & Gibbons, F. X. (2007). Social comparison: The end of a theory and the emergence of a field. Organizational Behavior and Human Decision Processes, 102(1), 3–21.

Calogero, R. M. (2012). Objectification theory, self-objectification, and body image. In T. Cash (Ed.), Encyclopedia of body image and human appearance (pp. 574–580). Academic Press. ISBN 978-0-12-384925-0.

Chen, J. K. K., Jones, S. W., & Heywood, A. J. (2011). Body dysmorphia, self-mutilation and the reconstructive surgeon. Journal of Plastic, Reconstructive & Aesthetic Surgery, 64(1), 4–8.

Daniels, E. A., Zurbriggen, E. L., & Ward, L. M. (2020). Becoming an object: A review of self-objectification in girls. Body Image, 33, 278–299.

Dijkslag, I. R., Bayens, L. R., Smeets, G., & Keuken, P. (2024). To beautify or uglify? The effects of Augmented Reality face filters on body satisfaction moderated by self-esteem and self-identification. Computers in Human Behavior, Article 108343.

du Rocher, A. R., Anderson, C.-A., Adhikar, Y., Leung, I., Lynch, H., Shah, M., Vincent, A., & Watkinson, K. (2020). Personality, self-appraisals, and body conscious emotions as predictors of symptoms of body dysmorphia and restrictive disordered eating. International Journal of Personality Psychology, 6, 27–36. https://doi.org/10.21827/ijpp.6.35666

Eshiet, J. (2020). Real me versus social media me: Filters, Snapchat dysmorphia, and beauty perceptions among young women. San Bernardino): Master's thesis, California State University.

Fahs, B., & Swank, E. (2017). Exploring stigma of "extreme" weight gain: The terror of fat possible selves in women's responses to hypothetically gaining one hundred pounds. Women's Studies International Forum, 61, 1–8. https://doi.org/10.1016/j.wsif.2016.11.004

Fardouly, J., Diedrichs, P. C., Vartanian, L. R., & Halliwell, E. (2015). Social comparisons on social media: The impact of Facebook on young women's body image concerns and mood. Body Image, 13, 38–45. https://doi.org/10.1016/j.bodyim.2014.12.002

Foroughi, A., Khanjani, S., & Moradi Asl, E. (2019). Relationship of concern about body dysmorphia with external shame, perfectionism, and negative affect: The mediating role of self-compassion. Iranian Journal of Psychiatry and Behavioral Sciences, 13(2). https://doi.org/10.5812/ijpbs.90186

Fredrickson, B. L., & Roberts, T. A. (1997). Objectification theory: Toward understanding women's lived experiences and mental health risks. Psychology of Women Quarterly, 21, 173–206.

Gibbons, F. X., & Buunk, B. P. (1999). Individual differences in social comparison: Development of a scale of social comparison orientation. Journal of Personality and Social Psychology, 76(1), 129.

Ho, S. S., Lee, E. W., & Liao, Y. (2016). Social network sites, friends, and celebrities: The roles of social comparison and celebrity involvement in adolescents' body image dissatisfaction. Social Media + Society, 2(3), Article 2056305116664216.

Javornik, A., Marder, B., Barhorst, J. B., McLean, G., Rogers, Y., Marshall, P., & Warlop, L. (2022). 'What lies behind the filter?' Uncovering the motivations for using augmented reality (AR) face filters on social media and their effect on well-being. Computers in Human Behavior, 128, Article 107126.

Jiang, S., & Ngien, A. (2020). The effects of Instagram use, social comparison, and self-esteem on social anxiety: A survey study in Singapore. Social Media+ Society, 6(2), 1–10. https://doi.org/10.1177/2056305120912488

Kleemans, M., Daalmans, S., Carbaat, I., & Anschütz, D. (2018). Picture perfect: The direct effect of manipulated Instagram photos on body image in adolescent girls. Media Psychology, 21(1), 93–110.

Lavrence, C., & Cambre, C. (2020). "Do I look like my selfie?": Filters and the digital-forensic gaze. Social Media + Society, 6(4). https://doi.org/10.1177/2056305120955182

Lewallen, J., & Behm-Morawitz, E. (2016). Pinterest or thinterest?: Social comparison and body image on social media. Social Media+ Society, 2(1). https://journals.sagepub.com/doi/abs/10.1177/2056305116640559

Mancin, P., Cerea, S., Bottesi, G., & Ghisi, M. (2024). Instagram use and negative and positive body image: The relationship with following accounts and content and filter use among female students. Current Psychology, 43(12), 10649–10661.

Matos, M., Coimbra, M., & Ferreira, C. (2023). When body dysmorphic symptomatology meets disordered eating: The role of shame and self-criticism. Appetite, 186, Article 106552. https://doi.org/10.1016/j.appet.2023.106552

McGovern, O., Collins, R., & Dunne, S. (2022). The associations between photo-editing and body concerns among females: A systematic review. Body Image, 43, 504–517.

Moradi, B., & Huang, Y. P. (2008). Objectification theory and psychology of women: A decade of advances and future directions. Psychology of Women Quarterly, 32(4), 377–398.

Morrachi, A., & Aaron, N. (2024). Social media beauty filters impact on mental health. NBC 24 News. https://nbc24.com/news/spotlight-on-america/social-media-beauty-filters-impacting-the-mental-health-of-young-women-tiktok-meta-snapchat-[illegible]

Myers, T. A., & Crowther, J. H. (2009). Social comparison as a predictor of body dissatisfaction: A meta-analytic review. Journal of Abnormal Psychology, 118(4), 683–698. https://doi.org/10.1037/a0016763

Pennell, H., & Behm-Morawitz, E. (2015). The empowering (super) heroine? The effects of sexualized female characters in superhero films on women. Sex Roles, 72(5–6), 211–220. https://doi.org/10.1007/s11199-015-0455-3

Phillips, K. A. (1998). Body dysmorphic disorder: Clinical aspects and treatment strategies. Bulletin of the Menninger Clinic, 62(4).

Raj, R., Anandpwar, A., Dwivedi, D., Pinnthu, S. P. M., Basu, P., & Agnihotri, B. K. (2022). Body dysmorphia and social media impact. International Journal of Health Sciences, 3725–3735.

Rizwan, B., Zaki, M., Javaid, S., Jaleem, Z., Mehmood, M., Riaz, M., ... Omar, H. (2022). Increase in body dysmorphia and eating disorders among adolescents due to social media: Increase in Body Dysmorphia and Eating Disorders Among Adolescents. Pakistan BioMedical Journal, 148–152.

Schmidt, A. M., Kapsis-Maddux, K., & Miller, S. A. (2010). Men's looking at your self-objectification, body image disturbance, and sorority rush. Sex Roles, 63, 6–17.

Ryan-Mosley. (2021). Beauty filters are changing the way young girls see themselves. MIT Technology Review. https://www.technologyreview.com/2021/04/02/1021635/beauty-filters-young-girls-augmented-reality-social-media/.

Saunders, J. F., & Eaton, A. A. (2018). Snaps, selfies, and shares: How three popular social media platforms contribute to the sociocultural model of disordered eating among young women. Cyberpsychology, Behavior, and Social Networking, 21(6), 343–354.

Schwartz, M. B., Vartanian, L. R., Nosek, B. A., & Brownell, K. D. (2006). The influence of one's own body weight on implicit and explicit anti-fat bias. Obesity, 14(3), 440–447. https://doi.org/10.1038/oby.2006.58

Seekis, V., Bradley, G. L., & Duffy, A. L. (2020). Appearance-related social networking sites and body image in young women: Testing an objectification-social comparison model. Psychology of Women Quarterly, 44(3), 377–392.

Sun, Q. (2021). Selfie editing and consideration of cosmetic surgery among young Chinese women: The role of self-objectification and facial dissatisfaction. Sex Roles, 84(11), 670–679.

Tiggemann, M., & Anderberg, I. (2020). Social media is not real: The effect of 'Instagram vs reality' images on women's social comparison and body image. New Media & Society, 22(12), 2183–2199. https://doi.org/10.1177/1461444819888720

Tiggemann, M., & McGill, B. (2004). The role of social comparison in the effect of magazine advertisements on women's mood and body dissatisfaction. Journal of Social and Clinical Psychology, 23(1), 23–44.

Vandenbosch, L., Fardouly, J., & Tiggemann, M. (2022). Social media and body image: Recent trends and future directions. Current Opinion in Psychology, 45, Article 101289. https://doi.org/10.1016/j.copsyc.2021.12.002

Verduyn, P., Gugushvili, N., Massar, K., Täht, K., & Kross, E. (2020). Social comparison on social networking sites. *Current Opinion in Psychology, 36*, 32–37. https://doi.org/10.1016/j.copsyc.2020.04.002

Verrastro, V., Liga, F., Cuzzocrea, F., & Gugliandolo, M. C. (2020). Fear the Instagram: beauty stereotypes, body image and Instagram use in a sample of male and female adolescents. *Interdisciplinary Journal of Technology, Culture and Education, 15*(1), 31–49.

Vogel, E. A., Rose, J. P., Roberts, L. R., & Eckles, K. (2014). Social comparison, social media, and self-esteem. *Psychology of Popular Media Culture, 3*(4), 206.

Body Image 26 (2018) 90–97

Contents lists available at ScienceDirect

Body Image

journal homepage: www.elsevier.com/locate/bodyimage

The effect of Instagram "likes" on women's social comparison and body dissatisfaction

Marika Tiggemann[a,*], Susannah Hayden[a], Zoe Brown[a], Jolanda Veldhuis[b]

[a] School of Psychology, Flinders University, Australia
[b] Department of Communication Science, VU University Amsterdam, Netherlands

ARTICLE INFO

Article history:
Received 6 April 2018
Received in revised form 10 July 2018
Accepted 10 July 2018
Available online 21 July 2018

Keywords:
Body image
Social networking sites
Instagram
Number of likes
Facial dissatisfaction

ABSTRACT

Photo-based activity on social networking sites has recently been identified as contributing to body image concerns. The present study aimed to investigate experimentally the effect of number of likes accompanying Instagram images on women's own body dissatisfaction. Participants were 220 female undergraduate students who were randomly assigned to view a set of thin-ideal or average images paired with a low or high number of likes presented in an Instagram frame. Results showed that exposure to thin-ideal images led to greater body and facial dissatisfaction than average images. While the number of likes had no effect on body dissatisfaction or appearance comparison, it had a positive effect on facial dissatisfaction. These effects were not moderated by Instagram involvement, but greater investment in Instagram likes w[...] more appearance comparison and facial dissatisfaction. The results illu[...] interactional aspects of social media (e.g., likes) can affect body image.

1. Introduction

Extensive research [...] of exposure to thin-i[...] or on television for th[...]ing of adolescent and y[...] Grabe, Ward, & Hyde, 20[...] 2009). More recent resea[...]net, with a particular fo[...] sites, such as Facebook. I[...]tics suggest that approxin[...] use social networking site[...] so on an at least daily basi[...] to create personal online [...]tion, and to form relationsh[...] same website. In contrast to [...] and television, social networ[...] [...]-generated (although it does also contai[...] [...]advertising), such that users are simultaneously information sources and receivers (Holland & Tiggemann, 2016). In addition, individuals are able to actively decide how, when, and for how long they wish to participate.

[...]ll but growing body of research has addressed the impact [...] networking sites, most commonly Facebook, on body [...] disordered eating outcomes. In their recent systematic [...] this research, Holland and Tiggemann (2016) concluded [...] increased social networking use is linked to body image [...] concerns, it is photo-based activity, e.g., posting photos [...] or making comments on others' photos (Meier & Gray, [...]) particularly salient. The authors also concluded that [...] evidence is largely correlational in design and called for [...]dinal and experimental approaches to determine the [...] of effects. The latter call reinforces Perloff's (2014) conclusion that there has been little experimental [...]dy image and newer media formats.

[...]ngly popular photo-based social networking site is [...] Instagram is a unique platform in that it is purely dedicated to the posting and sharing of photos, either with friends (on a private profile) or the wider public (on a public profile). Around 200 million people use Instagram on a daily basis (Statista, 2017). Instagram users are able to carefully select the personal photos they wish to post and to enhance them with Instagram filtering and editing tools in order to manage their self-presentation (Dumas, Maxwell-Smith, Davis, & Giulietti, 2017). Recent correlational research has shown that Instagram use is related to a variety of body image concerns (Cohen, Newton-John, & Slater, 2017; Fardouly, Willburger, & Vartanian, 2017; Feltman & Szymanski, 2018; Hendrickse, Arpan, Clayton, & Ridgway, 2017). As of yet there is no longitudinal evi-

* Corresponding author at: School of Psychology, Flinders University, GPO Box 2100, Adelaide, South Australia, 5001, Australia.
E-mail address: Marika.Tiggemann@flinders.edu.au (M. Tiggemann).

https://doi.org/10.1016/j.bodyim.2018.07.002

Body Image 26 (2018) 90–97

Contents lists available at ScienceDirect

Body Image

journal homepage: www.elsevier.com/locate/bodyimage

The effect of Instagram "likes" on women's social comparison and body dissatisfaction

Marika Tiggemann[a,*], Susannah Hayden[a], Zoe Brown[a], Jolanda Veldhuis[b]

[a] School of Psychology, Flinders University, Australia
[b] Department of Communication Science, VU University Amsterdam, Netherlands

ARTICLE INFO

Article history:
Received 6 April 2018
Received in revised form 10 July 2018
Accepted 10 July 2018
Available online 21 July 2018

Keywords:
Body image
Social networking sites
Instagram
Number of likes
Facial dissatisfaction

ABSTRACT

Photo-based activity on social networking sites has recently been identified as contributing to body image concerns. The present study aimed to investigate experimentally the effect of number of likes accompanying Instagram images on women's own body dissatisfaction. Participants were 220 female undergraduate students who were randomly assigned to view a set of thin-ideal or average images paired with a low or high number of likes presented in an Instagram frame. Results showed that exposure to thin-ideal images led to greater body and facial dissatisfaction than average images. While the number of likes had no effect on body dissatisfaction or appearance comparison, it had a positive effect on facial dissatisfaction. These effects were not moderated by Instagram involvement, but greater investment in Instagram likes was associated with more appearance comparison and facial dissatisfaction. The results illustrate how the uniquely social interactional aspects of social media (e.g., likes) can affect body image.

1. Introduction

Extensive research literature has documented negative effects of exposure to thin- ideal media images presented in magazines or on television for the body dissatisfaction and disordered eating of adolescent and young adult women (for meta-analyses, see Grabe, Ward, & Hyde, 2008; Groesz, Levine, & Murnen, 2002; Want, 2009). More recent research attention has shifted toward the Internet, with a particular focus on the impact of social networking sites, such as Facebook, Instagram, and Twitter. Australian statistics suggest that approximately 79% of adults (aged over 18 years) use social networking sites, with 89% of 18- to 29-year-olds doing so on an at least daily basis (Sensis, 2017). These sites allow users to create personal online profiles, to share photos and information, and to form relationships and interact with other users of the same website. In contrast to traditional media such as magazines and television, social networking content is largely peer-generated (although it does also contain some advertising), such that users are simultaneously information sources and receivers (Holland & Tiggemann, 2016). In addition, individuals are able to actively decide how, when, and for how long they wish to participate.

A small but growing body of research has addressed the impact of social networking sites, most commonly Facebook, on body image and disordered eating outcomes. In their recent systematic review of this research, Holland and Tiggemann (2016) concluded that while increased social networking use is linked to body image and eating concerns, it is photo-based activity, e.g., posting photos and viewing or making comments on others' photos (Meier & Gray, 2014), that is particularly salient. The authors also concluded that the existing evidence is largely correlational in design and called for more longitudinal and experimental approaches to determine the directionality of effects. The latter call reinforces Perloff's (2014) earlier general conclusion that there has been little experimental research on body image and newer media formats.

One increasingly popular photo-based social networking site is Instagram. Instagram is a unique platform in that it is purely dedicated to the posting and sharing of photos, either with friends (on a private profile) or the wider public (on a public profile). Around 200 million people use Instagram on a daily basis (Statista, 2017). Instagram users are able to carefully select the personal photos they wish to post and to enhance them with Instagram filtering and editing tools in order to manage their self-presentation (Dumas, Maxwell-Smith, Davis, & Giulietti, 2017). Recent correlational research has shown that Instagram use is related to a variety of body image concerns (Cohen, Newton-John, & Slater, 2017; Fardouly, Willburger, & Vartanian, 2017; Feltman & Szymanski, 2018; Hendrickse, Arpan, Clayton, & Ridgway, 2017). As of yet there is no longitudinal evi-

* Corresponding author at: School of Psychology, Flinders University, GPO Box 2100, Adelaide, South Australia, 5001, Australia.
E-mail address: Marika.Tiggemann@flinders.edu.au (M. Tiggemann).

https://doi.org/10.1016/j.bodyim.2018.07.002

dence, but initial experimental research has also shown that acute exposure to idealized Instagram images (attractive peers, celebrities) has a detrimental impact on body image (Brown & Tiggemann, 2016; Tiggemann & Zaccardo, 2015).

The negative effects of media exposure have generally been attributed to the process of social comparison (Levine & Murnen, 2009; Want, 2009). Social comparison theory (Festinger, 1954) argues that women evaluate their own appearance by comparing themselves with the sociocultural thin ideals of beauty presented in the media. Almost always this will constitute an upward comparison by which women fall short, resulting in dissatisfaction with their own body and appearance (Strahan, Wilson, Cressman, & Buote, 2006; Want, 2009). Furthermore, those individuals who are already anxious or uncertain about their body image seem to be particularly likely to seek out standards for (upward) social comparison, resulting in further body dissatisfaction (Want, 2009). In this, social comparison may be particularly pertinent to social networking sites. First, the ease with which individuals can connect to their networks at any time of the day gives rise to the opportunity for very fast and numerous comparisons on the basis of appearance (Tiggemann & Miller, 2010). Second, according to social comparison theory (Festinger, 1954), the drive for self-evaluation causes people to seek out comparisons with others who are similar rather than dissimilar to themselves. Thus peers, the major source of material on social media, provide more important comparison targets than do the models or celebrities featured in traditional media (Heinberg & Thompson, 1995). In support of this reasoning, Fardouly and Vartanian (2015) found that trait appearance comparison mediated the relationship between frequency of Facebook use and body image concerns, while Fardouly et al. (2017); Hendrickse et al. (2017), and Feltman and Szymanski (2018) showed similar relationships for Instagram use. More tellingly, in the Instagram experimental studies, the observed effects of idealized images on body dissatisfaction were shown to be mediated by state appearance comparison (Brown & Tiggemann, 2016; Tiggemann & Zaccardo, 2015).

Despite the fact that it is the interactivity of social media that most clearly distinguishes social media from traditional media (Perloff, 2014), as yet there has been little formal research on body image and the 'social' aspects of social media. The present study sought to begin this investigation by experimentally investigating the impact of one simple component, namely the number of likes on an Instagram image. Likes are an important and integral aspect of Instagram. Users can comment on and "like" photos, with "liking" being a very frequent activity (Frison & Eggermont, 2017). Importantly, the number of people who have "liked" a photo is then displayed under each image on Instagram for all to see. In this way, the number of likes can serve as a form of peer influence or social reinforcement. Social reinforcement theory postulates that the comments or actions of significant social agents, including media and peers, will reinforce particular attitudes and behaviours (Thompson & Stice, 2001). In the off-line environment, research has demonstrated that women and girls internalize the weight and shape ideals of their peer networks and share the resulting levels of body dissatisfaction (Lieberman, Gauvin, Bukowski, & White, 2001; Shroff & Thompson, 2006; Thompson & Stice, 2001). In the present case, a high number of likes attached to idealized (thin and attractive) Instagram photos should socially reinforce the importance of the beauty ideals displayed.

Initial qualitative research has indicated that likes are a marker of peer status and popularity (Dumas et al., 2017), as well as an accepted numerical indicator of consensually determined physical beauty (Chua & Chang, 2016). Indeed, Dumas et al. (2017) have documented a number of strategies that women use to actively seek likes for their own photos in order to obtain attention and validation, such as using filters or uploading photos at a certain time of day. Further, the adolescent girls in Chua and Chang's (2016) interview study viewed the number of likes that they receive as direct evaluative feedback about both their beauty and their self-worth. The number of likes on other people's photos (which Instagram displays) can also have consequences. For example, adolescents are significantly more likely to themselves "like" a photo if that photo has received more likes from peers (Jong & Drummond, 2016; Sherman, Payton, Hernandez, Greenfield, & Dapretto, 2016).

In the present study, we were interested in the effect of number of likes attached to others' photographs on women's own body image. We reasoned that the number of likes received by a photo would be taken as a reflection of the collective opinion of other Instagram users as to the worth and attractiveness of that person, serving to endorse that image and evoking greater attention and appearance comparison, and hence resulting in greater body dissatisfaction. Indeed, adolescent girls report that they compare themselves and their appearance more to peers with a higher number of likes on social media (Chua & Chang, 2016). In addition, the number of likes itself provides another attribute on which users can make social comparisons. Chua and Chang (2016) suggest that (upward) social comparison with Instagram images that receive a greater number of likes than what the individual would normally receive can lead to decreases in perceived self-worth and body satisfaction, while (downward) comparisons with Instagram users who receive fewer likes than they normally do may preserve body satisfaction and self-esteem. Although it is possible that habitual users of Instagram would be inured to the effect of likes, the importance placed on likes illustrated above led us to further reason that the effect of number of likes on body dissatisfaction might be greater for regular Instagram users who are familiar with the platform and who are more invested themselves in the number of likes they receive. These participants may be more inclined to attend to, interpret, and react to the number of likes on viewed images.

Thus, the main aim of the present study was to experimentally examine the effect of number of likes (low/high) on Instagram images (both thin-ideal and average figures) on women's body dissatisfaction. In addition, because faces may be particularly relevant in the social media context with the rise in the posting of self-portrait photos ("selfies"; Cohen, Newton-John, & Slater, 2018), we also included a simple measure of facial dissatisfaction. Based on the existing qualitative and correlational research, we predicted that viewing a greater number of likes would negatively affect body image by arousing greater social comparison on the basis of appearance and number of likes. In addition, we predicted that the effect of likes might be moderated by Instagram involvement, such that habitual users of Instagram or those who are invested in the feedback they receive would experience relatively greater body dissatisfaction in response to viewing images with a high number of likes.

2. Method

2.1. Design

The study employed a 2 (likes condition: low, high) × 2 (image type: thin-ideal, average) between-subjects experimental design. The major dependent variables were body dissatisfaction, facial dissatisfaction, and social comparison. Instagram use and investment were tested as potential moderating variables.

2.2. Participants

Participants were 220 female undergraduate students at Flinders University aged between 18 and 30 years. The major-

ity identified as Caucasian/White (67.3%), with 23.2% Asian, 2.3% African, 1.4% Aboriginal or Torres Strait Islander, and 5.9% 'other.' Participants were randomly allocated to one of the four experimental conditions of the design (subject to equal *n*), resulting in 55 participants per condition.

2.3. Materials

2.3.1. *Experimental manipulation: image type*

Two sets of stimulus materials were constructed for the study, each containing 15 Instagram images of thin-ideal or average women, respectively. All images were of Caucasian women, like the majority of the sample, and were sourced from public Instagram profiles using the hashtags "#fashion," "#beach," and "#plussize" to retrieve a range of body sizes. Images ranged from full body to head-and-shoulder shots, as these are common types of photographs found on Instagram, and were matched on clothing and degree of body shown.

The final images were selected from an initial pool of 120 Instagram images on the basis of ratings of thinness (1 = *extremely thin*, 5 = *extremely overweight*), attractiveness (1 = *extremely attractive*, 5 = *extremely unattractive*), and visual quality (1 = *excellent*, 5 = *very poor*) made by five independent raters in the target age range (M = 21.00 years, SD = 0.71). Images with a modal rating of '1' or '2' on thinness were considered as representative of the thin ideal, and images with rating of '3' or '4' were designated average. Any image rated a '5' by even a single rater was discarded to assist in the development of an average (rather than extremely overweight or unattractive) image set. An independent samples *t*-test confirmed that the final set of thin-ideal images were rated as significantly thinner (M = 1.93, SD = 0.26) than the average images (M = 3.47, SD = 0.52), $t(28)$ = 10.29, $p < .001$. As would be expected given that thinness is a defining feature of contemporary beauty, the thin-ideal images were also rated as significantly more attractive (M = 2.00, SD = 0.53) than the average images (M = 3.13, SD = 0.35), $t(28)$ = 6.86, $p < .001$. The thin-ideal and average image sets did not differ on visual quality (Ms = 2.53, 2.67, SDs = 0.52, 0.49), $t(28)$ = 0.73, p = .470.

2.3.2. *Experimental manipulation: number of likes*

Each image was presented within the Instagram frame with the Instagram logo, default profile picture icon, and a mock profile name (e.g., carol.t, Bree24) above the photo, and a specified number of Instagram likes below each image. In the Low Likes condition, the number of likes varied between 1 and 10. In the High Likes condition, the number of likes ranged between 100 and 300. These allocations were based on the pilot group's ratings of what they believed constituted a "low" (M = 12, SD = 4.47) and "high" number of likes (M = 300, SD = 122.47) for Instagram pictures. All other aspects of the presentation (e.g., profile name) remained constant across conditions.

2.3.3. *Social networking use*

Participants were provided with a list of social networking sites and asked which they used. For Facebook and Instagram, they were also asked how much time they spend there per day (*less than 10 min*, *10–30 min*, *30–60 min*, *1–2 hours*, *more than 2 h*), and the number of "friends" (Facebook), followers (Instagram), and people they follow (Instagram). Finally, they rated how much importance they placed on the visual quality of images on social networking sites (1 = *not important*, 5 = *very important*).

2.3.4. *Body and facial dissatisfaction*

Following Heinberg and Thompson (1995), visual analogue scales (VAS) were used to obtain measures of mood, body dissatisfaction, and facial dissatisfaction before and after viewing the Instagram images. The five mood items (not analysed here) were included to decrease the focus on appearance. Each scale consisted of a 100-mm horizontal line, with endpoints labelled *none* and *very much*. Participants were instructed to indicate how they were feeling "right now" by placing a small vertical mark on the line, which was subsequently measured to the nearest millimetre. As is normally done, the two body dissatisfaction dimensions ('weight dissatisfaction' and 'appearance dissatisfaction') were averaged to produce an overall body dissatisfaction score ranging from 0 to 100, with higher scores indicating greater body dissatisfaction. Because women also report high levels of facial dissatisfaction (Frederick, Kelly, Latner, Sandhu, & Tsong, 2016), a third VAS was added to specifically measure dissatisfaction with facial features, likely to be of increased relevance in the social media context (Cohen et al., 2018). In general, VAS carry the advantage that they can be completed quickly, are sensitive to small changes, and are difficult to recall in subsequent measurement. They have been shown to provide valid measures of body dissatisfaction, correlating significantly with longer and more complex measures of body image disturbance (Heinberg & Thompson, 1995). In the present sample, internal reliability for body dissatisfaction was acceptable at both pre-exposure (α = .85) and post-exposure (α = .90).

2.3.5. *State social comparison*

The amount of appearance comparison participants engaged in while viewing the images was assessed by Tiggemann and McGill's (2004) State Appearance Comparison Scale. The scale comprises three items which ask participants to rate on 7-point Likert-type scales the extent to which they thought about their appearance when viewing the Instagram photos (1 = *no thought about appearance*, 7 = *a lot of thought*), and the extent to which they compared their overall appearance and specific body parts, respectively, with those of the people they saw in the photos (1 = *no comparison*, 7 = *a lot of comparison*). The score for state appearance comparison was calculated by averaging the three items, producing a scale ranging from 1 to 7, with higher scores indicating greater state appearance comparison processing. Items in this scale are highly inter-correlated (r = .71 – .82) and the scale has been shown to be both reactive to experimental instruction and correlated with trait measures of physical appearance comparison (Tiggemann & McGill, 2004). In the present sample, the scale had good internal reliability (α = .92).

As the number of Instagram likes may themselves provide a further opportunity for social comparison, a fourth item was added. Participants were asked to rate the extent to which they compared the number of likes they personally receive with the number of likes on the viewed photographs (1 = *no comparison*, 7 = *a lot of comparison*).

2.3.6. *Instagram involvement*

Participants were asked a number of questions about their Instagram use. They were asked to indicate the number of pictures they upload to Instagram each month (*0*, *1–5*, *5–10*, *>10*), the average number of likes they receive on their posted photographs (*0–10*, *10–50*, *50–100*, *100–200*, *200+*), and the highest number of likes they have ever received. They were asked to rate the degree of importance that they place on the visual quality of photographs posted on Instagram by themselves and by others on 5-point Likert-type scales (1 = *not important*, 5 = *very important*). The final two items asked the degree of importance they placed on the number of likes on their own and someone else's photographs, respectively (1 = *not important*, 5 = *very important*). These latter two items were averaged to produce a measure of investment in likes with acceptable internal reliability (α = .76).

2.3.7. Recall

At the end of the questionnaire, participants were asked to indicate how many likes, on average, the photos they viewed had received (*0-20, 20–100, 100–200, 200-300*). This question was included to check that participants had noticed the number of likes attached to the images they viewed. They were also asked to rate the average thinness of the women in the images (1 = *not at all thin*, 7 = *very thin*) to check the assignment of images to experimental condition.

2.4. Procedure

Participants were recruited for a study entitled "Recreational Use of Instagram" and were tested individually or in small groups of two or three in the Psychology and Media research laboratory. After reading the Letter of Introduction and providing consent, participants completed the social networking use measure and the pre-exposure VAS measures of mood, body dissatisfaction, and facial dissatisfaction. Next, participants were presented with an iPad and viewed a slideshow of one of the four sets of experimental Instagram images (high/low likes × thin/average figures). Each image was displayed for 15 s. To ensure participants attended to the images, they were asked to rate the overall visual quality (e.g., blurriness, composition) of each image on a 5-point Likert-type scale (1 = *very poor*, 5 = *excellent quality*). Following exposure to the Instagram images, participants completed the post-exposure VAS, as well as measures of state appearance and likes comparison. Finally, participants completed the recall and Instagram use measures, before having their height and weight measured (with their consent). Testing sessions lasted approximately 30 min. Participants received course credit for their participation and were debriefed via an online system following completion of data collection. This protocol had received approval from the Institutional Research Ethics Committee.

3. Results

3.1. Characteristics of the sample

The women in the sample had a mean age of 20.13 years (SD = 2.58). Their mean body mass index (BMI) was 23.40 (SD = 4.73).

The social networking sites of Twitter, Pinterest, and Tumblr were each used by over a quarter of the women (25.5%, 39.5%, and 29.1% respectively). However, many more (nearly all) reported using Facebook (97.3%) and Instagram (93.2%). The modal use of both Facebook and Instagram was 30–60 min each day. Facebook users had an average of 649.4 (SD = 559.6) Facebook friends, while Instagram users had on average 516.7 (*SD* = 753.5) followers and themselves followed 386.4 (SD = 296.3) people. The latter figures are similar but a little lower than those of an equivalent United States sample of college students (579 followers and 493 people followed: Barry, Doucette, Loflin, Rivera-Hudson, & Herrington, 2017).

For Instagram specifically, the modal number of pictures uploaded every month was 1–5. The modal average number of likes received was 10–50, and the mean number of highest likes ever received was 146.15, but with wide variation (SD = 248.0). Participants placed considerable importance on the quality of the photographs posted on Instagram by others (M = 3.53, SD = 0.92) and especially by themselves (M = 4.20, SD = 0.81). Likewise, they placed moderate importance on the number of likes on their own (M = 3.17, SD = 1.18) and others' (M = 2.38, SD = 1.07) Instagram photos.

A series of one-way ANOVAs showed that the four experimental groups did not differ in age, $F(3, 216) = 1.76$, $p = .156$, $\eta_p^2 = .024$, BMI, $F(3, 216) = 1.10$, $p = .835$, $\eta_p^2 = .015$, or time spent on Instagram, $F(3, 215) = 0.91$, $p = .439$, $\eta_p^2 = .012$. They also did not differ on initial level of body dissatisfaction, $F(3, 216) = 0.84$, $p = .475$, $\eta_p^2 = .011$, or facial dissatisfaction, $F(3, 216) = 0.16$, $p = .924$, $\eta_p^2 = .002$, confirming that random assignment to experimental condition was successful.

3.2. Manipulation check

Participants in the high likes condition reported viewing significantly more likes (M = 2.01, SD = 0.52) than those in the low likes condition (M = 0.14, SD = 0.42), $t(218) = 29.78$, $p < .001$. These means are equivalent to 100–200 and 0–10 likes respectively, and so indicate that participants did attend to the number of likes on the images. Participants in the thin-ideal condition also rated the images they saw as significantly thinner (M = 5.72, SD = 1.04) than did participants in the average condition (M = 3.32, SD = 0.85), $t(218) = 18.77$, $p < .001$, confirming assignment of images to experimental condition.

3.3. Effect of likes on body and facial dissatisfaction

A 2 (likes condition: low, high) × 2 (image condition: thin-ideal, average) ANCOVA, with pre-exposure scores on body dissatisfaction entered as a covariate to control for individual differences, was conducted to test the effects of number of likes accompanying Instagram images on body dissatisfaction. The resulting adjusted means are displayed in Table 1. In contrast to prediction, the main effect of likes was not significant, $F(1, 215) = 0.18$, $p = .677$, $\eta_p^2 = .001$. The number of likes on viewed images did not result in greater body dissatisfaction. Nor was there any significant interaction with image type, $F(1, 215) = 0.19$, $p = .667$, $\eta_p^2 = .001$. There was, however, a significant main effect of image type, $F(1, 215) = 16.37$, $p < .001$, $\eta_p^2 = .071$, whereby the thin-ideal images led to greater body dissatisfaction than the average images.

The ANCOVA for facial dissatisfaction (pre-exposure facial dissatisfaction entered as a covariate) showed a significant main effect for number of likes, $F(1, 215) = 4.89$, $p = .028$, $\eta_p^2 = .022$. The adjusted means (Table 1) indicate that these lay in the opposite to predicted direction. Viewing a high number of likes led to lower, rather than greater, facial dissatisfaction. Again, the effect of image

Table 1
Means (SD) for Body Dissatisfaction, Facial Dissatisfaction, Appearance Comparison, and "Likes" Comparison by Experimental Condition.

	Condition				
	Average Images		Thin-ideal Images		
	Low Likes	High Likes	Low Likes	High Likes	
Body Dissatisfaction [a]	42.58 (1.54)	42.61 (1.54)	49.51 (1.54)	48.20 (1.55)	I
Facial Dissatisfaction [a]	40.54 (1.78)	38.30 (1.79)	47.26 (1.79)	41.58 (1.79)	I, L
Appearance Comparison	3.36 (1.47)	3.45 (1.56)	4.34 (1.40)	4.39 (1.65)	I
"Likes" Comparison	3.22 (1.97)	2.35 (1.73)	2.53 (1.82)	3.16 (1.77)	I x L

Note. I = significant effect (p < .05) of Image type; L = significant effect (p < .05) of Likes condition; I x L = significant interaction (p < .05).
[a] Adjusted means (SE).

type was significant, $F(1, 215) = 7.86$, $p = .006$, $\eta_p^2 = .035$ (exposure to thin-ideal images led to greater facial dissatisfaction than average images), with no significant interaction, $F(1, 215) = 0.92$, $p = .338$, $\eta_p^2 = .004$.

3.4. *Effect of likes on social comparison*

A 2 (likes condition: low, high) × 2 (image condition: thin-ideal, average) ANOVA showed no main effect of likes condition, $F(1, 216) = 0.02$, $p = .898$, $\eta_p^2 < .001$, nor interaction, $F(1, 216) = 0.15$, $p = .700$, $\eta_p^2 = .001$, on state appearance comparison. However, there was a significant main effect of image type, $F(1, 216) = 16.62$, $p < .001$, $\eta_p^2 = .071$, whereby the thin-ideal images aroused more state appearance comparison than the average images.

For likes comparison, although the mean score of 2.82 ($SD = 1.88$) suggests only a small-to-moderate amount of comparison on average on the basis of likes, scores ranged the full gamut of the scale (from 1 to 7) and the relatively large standard deviation indicates that some participants engaged in a great deal of likes comparison. A 2 (likes condition: low, high) × 2 (image condition: thin-ideal, average) ANOVA revealed no main effect of likes condition, $F(1, 216) = 0.22$, $p = .637$, $\eta_p^2 = .001$, or image type, $F(1, 216) = 0.07$, $p = .799$, $\eta_p^2 < .001$, on likes comparison. There was, however, a significant interaction between the number of likes and image type, $F(1, 216) = 9.13$, $p = .003$, $\eta_p^2 = .041$. As can be seen in Table 1, participants in the high likes condition reported a greater amount of likes comparison for the thin-ideal figure than those in the low likes condition. The opposite pattern was observed for the average figure.

3.5. *Role of comparison*

A series of hierarchical regressions was conducted to test whether social comparison (on the basis of appearance or number of likes) predicted increase in body dissatisfaction or facial dissatisfaction across experimental conditions. Pre-exposure body/facial dissatisfaction was entered on Step 1, followed by appearance or likes comparison on Step 2.

For state appearance comparison, it was found that Step 2 explained a significant amount of additional variance in post-exposure body dissatisfaction over and above initial body dissatisfaction, $\beta = .07$, $F_{change}(1, 217) = 4.61$, $p = .033$. Interestingly, when the high and low likes conditions were examined separately, there was significant prediction in the low likes condition, $\beta = .14$, $F_{change}(1, 107) = 10.88$, $p < .001$, but this prediction did not hold up in the high likes condition, $\beta = .01$, $F_{change}(1, 107) = 0.03$, $p = .856$. A similar pattern emerged for facial dissatisfaction. Although Step 2 failed to reach significance overall, $\beta = .06$, $F_{change}(1, 217) = 2.58$, $p = .109$, there was significant prediction in the low likes condition, $\beta = .12$, $F_{change}(1, 107) = 5.24$, $p = .024$, but not for a high number of likes, $\beta = .00$, $F_{change}(1, 107) = 0.00$, $p = .947$.

In contrast, comparison on the basis of likes did not offer significant prediction for either body dissatisfaction, $\beta = -.01$, $F_{change}(1, 217) = 0.08$, $p = .777$, or facial dissatisfaction, $\beta = -.04$, $F_{change}(1, 217) = 1.15$, $p = .285$. This remained the case when high and low likes conditions were examined separately.

3.6. *Moderating role of Instagram involvement*

A number of Instagram variables were conceptualized as potential moderators of the effect of number of likes on body and facial dissatisfaction. Specifically, it was proposed that the effect of likes on Instagram images would depend on the viewer's familiarity with Instagram and investment in likes. Table 2 displays the correlations between these potential moderators and outcome variables. It can be seen that each of time spent on Instagram, average number of likes, and likes investment was positively related to comparison on the basis of number of likes. Likes investment was also positively related to comparison on the basis of appearance and facial dissatisfaction before and after exposure.

To establish whether any of these moderated the effect of likes condition, a series of hierarchical multiple regressions was conducted. Where relevant, the pre-exposure measure was entered on Step 1, then likes condition (dichotomous variable) and the moderator variable in Step 2, followed by the two-way product term on the last step (Step 3). A significant interaction is established when the product term offers additional prediction beyond that provided by the prior variables (main effects). Here, the inclusion of the product term (Step 3) did not explain significant additional variance in any of body dissatisfaction, facial dissatisfaction, state appearance comparison, or likes comparison (all $R^2_{change} < .002$, all $p > .100$). Thus, none of the Instagram variables tested acted as moderators of the effect of number of likes.

4. Discussion

The major aim of the present study was to examine the effect of the number of likes accompanying Instagram images on body and facial dissatisfaction. In so doing, the study has extended the experimental investigation of social media and body image in general, and made a start on the important task of investigating the "social" aspects of social media. It was found that the number of likes had no effect on state appearance comparison or body dissatisfaction, but had a positive effect on facial dissatisfaction, whereby a high number of likes led to reduced facial dissatisfaction. This was the case for both thin-ideal and average images. Effects were not moderated by Instagram use or involvement, but women who reported greater investment in likes also showed more appearance comparison, likes comparison, and facial dissatisfaction.

It is clear that the sample as a whole is well-connected to social media with most participants using multiple platforms. Greater than 90% used Facebook and Instagram, with over a quarter also using other social media platforms, such as Twitter and Pinterest. Only two women (1%) in the present sample did not use either Facebook or Instagram. Data from equivalent samples (of the same age recruited from the same university) across time show that the rate of Instagram usage has steadily increased over recent years, from 78.5% (Tiggemann & Zaccardo, 2015) to 86.5% (Brown & Tiggemann, 2016) to the present 93.2%. This reflects the general trend in Australian figures, which also show that emerging adults, i.e., 18- to 29-year-olds (the present age group), are the most avid users of social media (Sensis, 2017).

Although not the major purpose of the study, the finding that exposure to thin-ideal Instagram images led to greater body and facial dissatisfaction relative to average images is consistent with a large body of previous experimental research demonstrating the negative effects of viewing thin idealized images (mostly drawn from fashion magazines) on women's body image (Grabe et al., 2008; Groesz et al., 2002; Want, 2009). It is also consistent with the results of two previous experimental studies using the Instagram platform (Brown & Tiggemann, 2016; Tiggemann & Zaccardo, 2015), but extends these by contrasting with average weight figures, rather than using nonhuman controls, the latter having been identified as a general confound in media studies (Ferguson, 2013). The size of effect on body dissatisfaction here was moderate-to-large, in contrast to the small effects typically reported in the earlier literature. As has been suggested (Tiggemann & Zaccardo, 2015), the size of effect may reflect the presentation of imagery via Instagram, an inherently photo-based platform and thus potentially a particularly potent form of transmission of ideals (Holland & Tiggemann, 2016; Meier & Gray, 2014). In addition, viewing the

Table 2
Correlations between Moderators and Outcome Variables.

	AppComp	Likes Comp	Pre BD	Post BD	Pre FD	Post FD
Instagram Use	.06	.19**	.07	.03	.01	.03
Average Likes	.07	.18**	.00	.04	-.01	.01
Likes Investment	.26***	.39***	.10	.04	.19**	.16*

Note: App Comp = Appearance comparison; Likes Comp = Likes comparison; Pre BD = pre-body dissatisfaction; Post BD = post-body dissatisfaction; Pre FD = pre-facial dissatisfaction; Post FD = post-facial dissatisfaction.
*$p < .05$; ** $p < .01$; *** $p < .001$.

images on a mobile device, in this case an iPad, provides a contemporary and ecologically valid medium that may add immediacy.

The first major finding here was that the number of likes did not affect body dissatisfaction. We found no support for the proposition that viewing Instagram images with a high number of Instagram likes would result in relatively greater body dissatisfaction. However, we also found that a high number of likes did not evoke greater appearance comparison as we had reasoned. Logically, then, body dissatisfaction would not be affected. This is surprising in light of Chua and Chang's (2016) finding that adolescent girls explicitly report comparing themselves more to peers on social media with a higher number of likes. It is possible that likes are particularly meaningful for adolescent girls who place so much importance on peer opinion, but no longer carry the same impact by early adulthood, the developmental stage of the present sample. It also needs to be recognised that the photos shown in the present experimental protocol were of unknown peers, whereas Chua and Chang's (2016) girls may have been responding to known peers in their networks. Relatedly, Chua and Chang's (2016) sample were Singaporean and there may be important cultural differences in which aspects of social media are most valued. Nevertheless, the number of likes was not irrelevant here. Only in the low likes condition did the amount of appearance comparison processing that women engaged in predict increase in body dissatisfaction as is usually found (e.g., Tiggemann & Polivy, 2010; Tiggemann, Polivy, & Hargreaves, 2009) and consistent with sociocultural models of body image (Thompson, Heinberg, Altabe, & Tantleff-Dunn, 1999; Tiggemann, 2011). It is not clear why this was not also the case in the high likes condition. Perhaps images with a high number of likes evoke other cognitive processes (e.g., appearance schemas, reinforcement of beauty ideals) or emotional states (e.g., envy) that were not measured but may be pertinent to body dissatisfaction. Future research will need to determine the predictors of body image change under these circumstances.

The second major finding was that the number of likes did have a significant effect on facial dissatisfaction. In particular, viewing a high number of likes led to lower dissatisfaction with one's own face. Although not predicted, this positive effect is perhaps consistent with a view of likes as an online form of supportive communication, identified by Oh, Ozkaya, and LaRose, (2014) as providing a sense of community, peer belonging, and life satisfaction. It is also possible that social comparison on the basis of facial features may not be as universally negative (upward) as comparison on the basis of body size and shape seem to be. Facial ideals are likely more complex, individual, and less narrowly prescribed than are body ideals. To test the applicability of sociocultural models (Thompson et al., 1999; Tiggemann, 2011) in the facial domain, future research might ascertain which facial features are considered ideal and measure the extent of comparison on the basis of facial features.

We further attempted to explore and explicitly measure (albeit with only a single item) a new type of comparison unique to social media, namely comparison on the basis of number of likes. The results indicated that some participants certainly engage in this type of comparison. Further, the experimental manipulations had an effect on this form of comparison. For the thin-ideal figure, viewing images with a high number of likes evoked greater comparison on the basis of likes than viewing images with low likes, presumably reflecting upward comparison, and consistent with Chua and Chang's (2016) qualitative descriptions. For the average figure, however, fewer likes evoked relatively greater (presumably downward) comparison. This interaction may be a function of matching participants' expectations, in that individuals may expect thin-ideal images to receive a large number of likes and average images to receive fewer likes. Regardless, the results show that the meaning of number of likes will differ depending on the content of the material to which they are attached. Although likes comparison did not predict change in body or facial dissatisfaction, it remains possible that it might affect other more global outcomes not measured here, e.g., self-esteem.

The third major finding was that, in contrast to our prediction that habitual and highly invested users of Instagram would be particularly reactive to the number of likes on an image, we observed no moderating effect on any outcome measure. Perhaps the relatively large effect of figure size on our dependent variables somehow overshadowed any other potential effects. Nevertheless, again these aspects were not irrelevant. In particular, we found that, regardless of experimental condition, the extent to which individuals placed importance on number of likes was positively related to greater comparison on both appearance and likes, as well as more facial dissatisfaction.

The difference in obtained findings between body and facial dissatisfaction illustrates a more general important point. When addressing newer media formats, we should be careful to align our research questions and associated protocols and measures to the new format, rather than simply and uncritically transferring existing protocols and measures from traditional media research. Women and girls reportedly spend a great deal of time and effort in taking, selecting and editing the photographs of themselves ("selfies") they choose to upload on Instagram (Chua & Chang, 2016; Dumas et al., 2017). As these are commonly close-ups, the face may be an increasingly important target feature for investigations of social media use. Future research might also be directed at developing psychometrically robust measures of activities unique to social media, such as comparison on the basis of number of likes.

Taken together, the present findings indicate both positive and negative aspects of likes attached to Instagram images. On the one hand, viewing images with a high number of likes led to greater satisfaction with women's own facial features, perhaps reflecting a positive sense of online social support and communication. On the other hand, being highly invested in likes was related to appearance comparison and facial dissatisfaction, in a way that general Instagram use and average number of likes were not. Thus, it appears that it is the importance placed on the number of likes received that may be particularly problematic, rather than the receiving and viewing of likes in general, which may be regarded as normative aspects of Instagram participation. This result carries the important practical implication that women and girls should be dissuaded from actively seeking likes on their Instagram images and, in particular, from viewing the number of likes they receive as an indication

of their beauty or self-worth. This specific message could well be added to current media literacy programs, which have shown some effectiveness in protecting body image (Levine & Murnen, 2009; Yager, Diedrichs, Ricciardelli, & Halliwell, 2013).

Like all studies, the present study carries a number of limitations. First, the sample consisted of Australian university students and thus results may not generalize to other groups of women. In particular, the number of likes may be more salient for adolescent girls who are also high users of social networking sites (Sensis, 2017). Second, the study took place in a laboratory setting. Although the viewed images were real images sourced from Instagram and presented in an ecologically valid format (on an iPad), participants were asked to attend to the images (by evaluating visual quality) in a way that they would not normally do in more naturalistic settings. Third, all viewed images were of Caucasian women to match the majority of the sample. Images representing greater racial and ethnic diversity have not yet been trialled in Instagram experimental studies. Fourth, the likes investigated here were on other people's photos. The significance and implications of number of likes would undoubtedly be much greater were they attached to the participant's own photos. Fifth, the images (and associated number of likes) were viewed passively. Unlike on social networking sites, participants had no opportunity to engage in interaction, for example, by also "liking" the image. Future research will need to devise creative methodologies to capture more fully the effects on body image of self-relevant posts and social networking behaviours.

Despite the above limitations, the present study has made a start in investigating one important and quintessential aspect of social media, namely the number of likes. Future research might continue this investigation by addressing other important aspects, such as the making or receiving of comments. The findings offer some insight into how online social interaction may contribute to body image in a way that is very different from traditional media formats like fashion magazines and television. In so doing, they illustrate the complexity and multi-faceted nature of the contemporary media world in which young adults live.

References

Barry, C., Doucette, H., Loflin, D., Rivera-Hudson, N., & Herrington, L. (2017). "Let me take a selfie": Associations between self-photography, narcissism, and self-esteem. *Psychology of Popular Media Culture*, *6*, 48–60. http://dx.doi.org/10.1037/ppm0000089

Brown, Z., & Tiggemann, M. (2016). Attractive celebrity and peer images on Instagram: Effect on women's mood and body image. *Body Image*, *19*, 37–43. http://dx.doi.org/10.1016/j.bodyim.2016.08.007

Chua, T., & Chang, L. (2016). Follow me and like my beautiful selfies: Singapore teenage girls' engagement in self-presentation and peer comparison on social media. *Computers in Human Behavior*, *55*, 190–197. http://dx.doi.org/10.1016/j.chb.2015.09.011

Cohen, R., Newton-John, T., & Slater, A. (2017). The relationship between Facebook and Instagram appearance-focused activities and body image concerns in young women. *Body Image*, *23*, 183–187. http://dx.doi.org/10.1016/j.bodyim.2017.10.002

Cohen, R., Newton-John, T., & Slater, A. (2018). 'Selfie'-objectification: The role of selfies in self-objectification and disordered eating in young women. *Computers in Human Behavior*, *79*, 68–74. http://dx.doi.org/10.1016/j.chb.2017.10.027

Dumas, T., Maxwell-Smith, M., Davis, J., & Giulietti, P. (2017). Lying or longing for likes? Narcissism, peer belonging, loneliness and normative versus deceptive like-seeking on Instagram in emerging adulthood. *Computers in Human Behavior*, *71*, 1–10. http://dx.doi.org/10.1016/j.chb.2017.01.037

Fardouly, J., & Vartanian, L. (2015). Negative comparisons about one's appearance mediate the relationship between Facebook usage and body image concerns. *Body Image*, *12*, 82–88. http://dx.doi.org/10.1016/j.bodyim.2014.10.004

Fardouly, J., Willburger, B., & Vartanian, L. (2017). Instagram use and young women's body image concerns and self-objectification: Testing mediational pathways. *New Media and Society*. http://dx.doi.org/10.1177/1461444817694499

Feltman, C., & Szymanski, D. (2018). Instagram use and self-objectification: The roles of internalization, comparison, appearance commentary, and feminism. *Sex Roles*, *78*, 311–324. http://dx.doi.org/10.1007/s11199-017-0796-1

Ferguson, C. (2013). In the eye of the beholder: Thin-ideal media affects some, but not most, viewers in a meta-analytic review of body dissatisfaction in women and men. *Psychology of Popular Media Culture*, *2*, 20–37. http://dx.doi.org/10.1037/a0030766

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*, 117–140. http://dx.doi.org/10.1177/001872675400700202

Frederick, D., Kelly, M., Latner, J., Sandhu, G., & Tsong, Y. (2016). Body image and face image in Asian American and white women: Examining associations with surveillance, construal of self, perfectionism, and sociocultural pressures. *Body Image*, *16*, 113–125. http://dx.doi.org/10.1016/j.bodyim.2015.12.002

Frison, E., & Eggermont, S. (2017). Browsing, posting, and liking on Instagram: The reciprocal relationships between different types of Instagram use and adolescents' depressed mood. *Cyberpsychology, Behavior, and Social Networking*, *20*, 603–609. http://dx.doi.org/10.1089/cyber.2017.0156

Grabe, S., Ward, L., & Hyde, J. (2008). The role of the media in body image concerns among women: A meta-analysis of experimental and correlational studies. *Psychological Bulletin*, *134*, 460–476. http://dx.doi.org/10.1037/0033-2909.134.3.460

Groesz, L., Levine, M., & Murnen, S. (2002). The effect of experimental presentation of thin media images on body satisfaction: A meta-analytic review. *International Journal of Eating Disorders*, *31*, 1–16. http://dx.doi.org/10.1002/eat.10005

Heinberg, L., & Thompson, J. (1995). Body image and televised images of thinness and attractiveness: A controlled laboratory investigation. *Journal of Social and Clinical Psychology*, *14*, 325–338. http://dx.doi.org/10.1521/jscp.1995.14.4.325

Hendrickse, J., Arpan, L., Clayton, R., & Ridgway, J. (2017). Instagram and college women's body image: Investigating the roles of appearance-related comparisons and intrasexual competition. *Computers in Human Behavior*, *74*, 92–100. http://dx.doi.org/10.1016/j.chb.2017.04.027

Holland, G., & Tiggemann, M. (2016). A systematic review of the impact of the use of social networking sites on body image and disordered eating outcomes. *Body Image*, *17*, 100–110. http://dx.doi.org/10.1016/j.bodyim.2016.02.008

Jong, S., & Drummond, M. (2016). Hurry up and 'like' me: Immediate feedback on social networking sites and the impact on adolescent girls. *Asia-Pacific Journal of Health, Sport and Physical Education*, *7*, 251–267. http://dx.doi.org/10.1080/18377122.2016.1222647

Lieberman, M., Gauvin, L., Bukowski, W. M., & White, D. (2001). Interpersonal influence and disordered eating behaviors in adolescent girls: The role of peer modeling, social reinforcement, and body-related teasing. *Eating Behaviors*, *2*, 215–236. http://dx.doi.org/10.1016/S1471-0153(01)00030-7

Levine, M., & Murnen, S. (2009). Everybody knows that mass media are/are not [pick one] a cause of eating disorders: A critical review of evidence for a causal link between media, negative body image, and disordered eating in females. *Journal of Social and Clinical Psychology*, *28*, 9–42. http://dx.doi.org/10.1521/jscp.2009.28.1.9

Meier, E., & Gray, J. (2014). Facebook photo activity associated with body image disturbance in adolescent girls. *Cyberpsychology, Behavior, and Social Networking*, *17*, 199–206. http://dx.doi.org/10.1089/cyber.2013.0305

Oh, H., Ozkaya, E., & LaRose, R. (2014). How does online social networking enhance life satisfaction? The relationships among online supportive interaction, affect, perceived social support, sense of community, and life satisfaction. *Computers in Human Behavior*, *30*, 69–78. http://dx.doi.org/10.1016/j.chb.2013.07.053

Perloff, R. (2014). Social media effects on young women's body image concerns: Theoretical perspectives and an agenda for research. *Sex Roles*, *71*, 363–377. http://dx.doi.org/10.1007/s11199-014-0384-6

Sensis. (2017). *Sensis social media report 2017* Retrieved from. https://www.sensis.com.au/asset/PDFdirectory/Sensis-Social-Media-Report-2017.pdf

Sherman, L., Payton, A., Hernandez, L., Greenfield, P., & Dapretto, M. (2016). The power of the like in adolescence. *Psychological Science*, *27*, 1027–1035. http://dx.doi.org/10.1177/0956797616645673

Shroff, H., & Thompson, J. (2006). Peer influences, body-image dissatisfaction, eating dysfunction and self-esteem in adolescent girls. 2018 dysfunction and self-esteem in adolescent girls. *Journal of Health Psychology*, *11*, 533–551. http://dx.doi.org/10.1177/1359105306065015

Statista. (2017). *Instagram daily active users 2017* Retrieved from. https://www.statista.com/statistics/657823/number-of-daily-active-instagram-users/

Strahan, E., Wilson, A., Cressman, K., & Buote, V. (2006). Comparing to perfection: How cultural norms for appearance affect social comparisons and self-image. *Body Image*, *3*, 211–227. http://dx.doi.org/10.1016/j.bodyim.2006.07.004

Thompson, J., & Stice, E. (2001). Thin-ideal internalization: Mounting evidence for a new risk factor for body-image disturbance and eating pathology. *Current Directions in Psychological Science*, *10*, 181–183. http://dx.doi.org/10.1111/1467-8721.00144

Thompson, J., Heinberg, L., Altabe, M., & Tantleff-Dunn, S. (1999). *Exacting beauty: Theory, assessment, and treatment of body image disturbance*. Washington, DC: American Psychological Association.

Tiggemann, M. (2011). Sociocultural perspectives on human appearance and body image. In T. F. Cash & L. Smolak (Eds.), *Body image: A handbook of science, practice, and prevention* (pp. 12–19). New York: Guilford.

Tiggemann, M., & McGill, B. (2004). The role of social comparison in the effect of magazine advertisements on women's mood and body dissatisfaction. *Journal of Social and Clinical Psychology*, *23*, 23–44. http://dx.doi.org/10.1521/jscp.23.1.23.26991

Tiggemann, M., & Miller, J. (2010). The internet and adolescent girls' weight satisfaction and drive for thinness. *Sex Roles*, *63*, 79–90. http://dx.doi.org/10.1007/s11199-010-9789-z

Tiggemann, M., & Polivy, J. (2010). Upward and downward: Social comparison processing of thin idealized media images. *Psychology of Women Quarterly*, 34, 356–364. http://dx.doi.org/10.1111/j.1471-6402.2010.01581.x

Tiggemann, M., Polivy, J., & Hargreaves, D. (2009). The processing of thin ideals in fashion magazines: A source of social comparison or fantasy? *Journal of Social and Clinical Psychology*, 28, 73–93. http://dx.doi.org/10.1521/jscp.2009.28.1.73

Tiggemann, M., & Zaccardo, M. (2015). "Exercise to be fit, not skinny": The effect of fitspiration imagery on women's body image. *Body Image*, 15, 61–67. http://dx.doi.org/10.1016/j.bodyim.2015.06.003

Want, S. (2009). Meta-analytic moderators of experimental exposure to media portrayals of women on female appearance satisfaction: Social comparisons as automatic processes. *Body Image*, 6, 257–269. http://dx.doi.org/10.1016/j.bodyim.2009.07.008

Yager, Z., Diedrichs, P., Ricciardelli, L., & Halliwell, E. (2013). What works in secondary schools? A systematic review of classroom-based body image programs. *Body Image*, 10, 271–281. http://dx.doi.org/10.1016/j.bodyim.2013.04.001

MEDIA PSYCHOLOGY

Routledge
Taylor & Francis Group

Media Psychology

ISSN: 1521-3269 (Print) 1532-785X (Online) Journal homepage: https://www.tandfonline.com/loi/hmep20

Picture Perfect: The Direct Effect of Manipulated Instagram Photos on Body Image in Adolescent Girls

Mariska Kleemans, Serena Daalmans, Ilana Carbaat & Doeschka Anschütz

To cite this article: Mariska Kleemans, Serena Daalmans, Ilana Carbaat & Doeschka Anschütz (2018) Picture Perfect: The Direct Effect of Manipulated Instagram Photos on Body Image in Adolescent Girls, Media Psychology, 21:1, 93-110, DOI: 10.1080/15213269.2016.1257392

To link to this article: https://doi.org/10.1080/15213269.2016.1257392

Beauty Filters

MEDIA PSYCHOLOGY
2018, VOL. 21, NO. 1, 93–110
https://doi.org/10.1080/15213269.2016.1257392

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Picture Perfect: The Direct Effect of Manipulated Instagram Photos on Body Image in Adolescent Girls

Mariska Kleemans, Serena Daalmans, Ilana Carbaat, and Doeschka Anschütz

Behavioural Science Institute, Radboud University, Nijmegen, The Netherlands

ABSTRACT

This study investigates the effect of manipulated Instagram photos on adolescent girls' body image, and whether social comparison tendency moderates this relation. A between-subject experiment was conducted in which 144 girls (14–18 years old) were randomly exposed to either original or manipulated (retouched and reshaped) Instagram selfies. Results showed that exposure to manipulated Instagram photos directly led to lower body image. Especially, girls with higher social comparison tendencies were negatively affected by exposure to the manipulated photos. Interestingly, the manipulated photos were rated more positively than the original photos. Although the use of filters and effects was detected, reshaping of the bodies was not noticed very well. Girls in both conditions reported to find the pictures realistic. Results of this study implied that the recent societal concern about the effects of manipulated photos in social media might be justified, especially for adolescent girls with a higher social comparison tendency.

Instagram is currently a very popular social network site, especially among teenagers (Seetharaman, 2015). Instagram allows its users to share photos and videos with others. Since its start in 2010, it has attracted more than 400 million active users, who upload around 80 million photos a day (Instagram, 2015). Photos and videos are a very direct form of online self-presentation and have become an increasingly powerful form of social online currency (Rainie, Brenner, & Purcell, 2012). Even though Instagram is the most popular photo sharing application on the Internet, it has received very little academic attention (Hu, Manikonda, & Kambhampati, 2014). This is surprising as Instagram has lately been a topic of concern in the public debate (Sass, 2014; Winter, 2013). The main concern involves the possibility to manipulate Instagram photos by using retouching techniques and, consequently, the potentially negative influence that these "perfect pictures" may have on body image of (young) Instagram users. In particular, both critics and fans frequently blame celebrities and models for using photo

CONTACT Serena Daalmans s.daalmans@maw.ru.nl Behavioural Science Institute, Radboud University, P. O. Box 9104, 6500 HE Nijmegen, The Netherlands.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/HMEP

enhancement and retouching techniques. Hence, they normalize an unrealistic body ideal, which is problematic as they serve as role models for girls and young women (Sullivan, 2014).

Although in general, famous people have been criticized for manipulating self-images on social media, there are important reasons to investigate the effects of edited pictures of "ordinary" Instagram users. Research has indicated that men and women, both adolescents and adults, compare themselves more often to peers than to models or celebrities for social attributes (i.e., personality, intelligence) and physical attributes (i.e., weight, height, body image; Jones, 2001; Strahan, Wilson, Cressman, & Buote, 2006), and has thereby supported the general expectation from the social comparison literature that individuals generally prefer to make social comparisons to similar others (Miller, Turnbull, & McFarland, 1988). Furthermore, the comparison with peers might affect their body image in a comparable manner as media images do (Myers & Crowther, 2009). This might be due to the fact that peers are perceived more similar to themselves than celebrities and therefore are more relevant to compare themselves with. This is in line with the extensive identification literature, combining social cognitive theory (Bandura, 2002), the message-interpretation process model (Austin & Meili, 1994), and exemplification theory (Zillmann & Brosius, 2000). Shortly summarized, these theories state that when people perceive others to be more similar to themselves, identification and related cognitive and behavioral consequences are more likely to occur (see also Andsager, Bemker, Choi, & Torwel, 2006). This social influence mechanism might just as well apply to social media networks, as these are very popular environments for peer interaction. Research revealed that users of social media platforms often manipulate their appearance in the pictures they post online, and that this habit is especially prevalent among young girls (Manago, Graham, Greenfield, & Salimkhan, 2008; Philly Renfrew Center Foundation, 2014). However, the effects of exposure to enhanced social media photos of peers on young girls' body image are still largely unknown. Adolescent girls are often found to be particularly vulnerable for being influenced by media images (e.g., Borzekowski, Robinson, & Killen, 2000) because of the psychosocial development that is characteristic for this phase (Sturdevant & Spear, 2002). The present study attempts to further elucidate this relation by investigating the effects of exposure to original and manipulated Instagram photos of peers on adolescent girls' body image.

Earlier research focusing on body image has primarily investigated the influence of exposure to idealized thin bodies in advertisements, magazines, television, as well as music videos on young women's body image. These studies often revealed a relation between exposure to the thin ideal and a negative body image among young girls and women (e.g., Grabe, Ward, & Hyde, 2008; Halliwell & Dittmar, 2004; Irving, 1990;). This effect can be explained by the negative contrast theory, stating that women experience a contrast between themselves and the thin, idealized models and that this

leads to lower body satisfaction (Thornton & Maurice, 1999; Thornton & Moore, 1993). However, some studies actually found self-enhancing effects of exposure to thin ideal images (e.g., Henderson-King & Henderson-King, 1997; Joshi, Herman, & Polivy, 2004; Mills, Polivy & Tiggemann, 2002; Myers & Biocca, 1992). Based on these findings, an alternative to the negative contrast theory was formulated by Mills et al. (2002), suggesting that thin media models might cause a "thinness fantasy" (Myers & Biocca, 1992) by inspiring women for whom thinness is self-relevant. Ample research has studied the effects of media models on body image, but the effects of exposure to images on social media sites are not well established. As traditional media are surpassed in popularity by online social media platforms, especially among young people, it becomes important to include these newer forms of media in this line of research as well (Fardouly & Vartanian, 2015; Seetharaman, 2015; Tiggemann & Slater, 2013).

One important characteristic that sets social media apart from other studied media types is the strong focus on peer interactions. Media models, that is, models and celebrities, are often presented as unrealistic standards of beauty in for example media literacy programs and the public debate, because of the well-known editing and retouching techniques used when displaying media models (e.g., Thompson & Heinberg, 1999; Yamamiya, Cash, Melnyk, Posavac, & Posavac, 2005). Less known is that "ordinary" social media users also use these techniques, as a part of impression management in self-presentation (Manago et al., 2008; Won Kim & Chock, 2015). Girls who compare themselves with manipulated photos of peers might think they are comparing themselves with people who are similar to them, rather than with celebrities whose bodies are seen as unattainable (Jones, 2001). However, one might conclude that the appearances of these peers might be not realistic at all. The current study investigates whether manipulated—and thereby idealized—Instagram photos of peers affect body image in young women. In line with earlier studies on exposure to idealized images, it is expected that:

H1: Exposure to manipulated Instagram photos leads to lower body satisfaction in adolescent girls than exposure to original photos.

Previous research also revealed that the effects of exposure to the thin ideal in traditional media depend on individual susceptibility factors (e.g., Henderson-King & Henderson-King, 1997; Joshi et al., 2004; Myers & Biocca, 1992; Wilcox & Laird, 2000). Especially the tendency to engage in social comparisons (Social Comparison Theory; Festinger, 1954) has proven to be influential in the relation between exposure to the thin ideal in the media and women's body dissatisfaction (e.g., Keery, Van den Berg, & Thompson, 2004; Won Kim & Chock, 2015).

It is often found that body dissatisfaction is a result of young women's upward social comparisons of their own appearance with the appearance of other young women in real life or in a (social) media context (e.g., Dittmar & Howard, 2004; Fardouly, Diedrichs, Vartanian, Halliwell, 2015; Fardouly & Vartanian, 2015; Mabe, Forney, & Keel, 2014; Tiggemann & Miller, 2010; Tiggemann & Slater, 2013; Vartanian & Dey, 2013). More precisely, women who more frequently engage in comparisons with others are also more negatively affected by exposure to idealized images of others, compared to women who engage in these comparisons less frequently (Dittmar & Howard, 2004). Therefore, the current study examines whether women's social comparison tendency moderates their responses to manipulated Instagram photos. It is expected that:

H2: The negative effect of exposure to manipulated Instagram photos compared to original Instagram photos on body satisfaction is stronger for girls with higher social comparison tendency.

Method

An online experiment was conducted to investigate the effect of manipulated Instagram photos on the body image of girls. The experiment has a 2 (Instagram photos: original versus manipulated) × 2 (social comparison tendency: lower vs. higher) between-subjects design. Participants were randomly exposed to either 10 original Instagram photos (N = 72) or to 10 manipulated photos (N = 72). Subsequently, they answered a number of questions through an online survey.

Participants and procedure

A total number of 144 adolescent girls participated in the experiment. Their age ranged between 14 and 18 years old (M = 15.92; SD = 1.16). The girls in our sample attended different levels of secondary education. In The Netherlands—where this study was conducted—children are divided over different educational levels at secondary schools based on their achievement scores obtained at elementary schools (cf. Scheerens, Luyten, & Van Ravens, 2011). Students can either attend pre-vocational secondary education (the lowest level, preparing for vocational education), general secondary education (middle level, preparing students for universities of applied sciences), or pre-university education (highest level, preparing student for research universities). As the transition from elementary to secondary school usually takes place at the age of 12, the girls in our sample attended at least of few years of secondary education at a specific level of secondary

education. This makes it important to take level of education into account. The girls in our sample were almost equally divided over the three levels of education that could be discerned: 49 girls attended a low level of education, 50 girls had a medium level of education, and 45 girls attended the highest level of education. The division of age over the different levels of education was somewhat skewed: The lower educated participants were somewhat younger (M = 15.18; SD = .81) than participants attending the middle (M = 16.22; SD = .93) or highest level (M = 16.40; SD = 1.32) of education. This can be explained by the fact that education at the lowest level takes 4 years to complete, whereas education at the middle (5 years) and highest level (6 years) takes longer. Therefore, students at the latter two levels can be older when attending secondary education. Preliminary analyses, however, showed that the small differences in the division of age over the three levels of education did not play a role of interest in the analyses.

Snowball sampling was used to recruit participants. We first invited girls from our own network to participate. Next, we asked them whether they knew other girls aged between 14 and 18 years old that might want to participate. Because most of the participants were under the age of 18, active parental consent was always asked for prior to the start of the data collection. This procedure is in accordance with the guidelines as formulated by the ethics committee of Faculty of Social Sciences (ECSS) at Radboud University. After obtaining permission from parents, girls received an email containing further information about the study and a link to the online experiment.

In the invitation email, a cover story was used to inform them about the research, as it was important not to reveal the real aim of the study. Participants were told that the study goal was to investigate how contextual factors affect preferences for different face types, and that they therefore would be exposed to pictures of people with different facial expressions. The email also contained instructions about the procedure they had to follow while taking part in the study. We asked them to complete the task at a moment that they were in a quiet area, without disturbing factors in their surroundings. In addition, they were asked to focus on the experiment only and to avoid interruptions.

After clicking on the link to start the experiment, a short instruction was presented on the screen. We repeated the (false) study aim and told them that the study started with showing them 10 Instagram photos, either original or manipulated. We asked each participant to take enough time to carefully look at the photos, and informed them that subsequently a number of questions would be asked. We emphasized that all information provided would be treated confidentially. After completing the study, participants were thanked for their voluntary participation. Moreover, they were offered the possibility to contact the researchers through email in case they wanted to have more information or to ask additional questions about the study.

Materials

The stimulus materials consisted of 10 "selfies" (self-portrait photos taken with a digital camera or camera phone held in the hand; Saltz, 2014). Selfies were used because photos in this format are a popular trend on Instagram and other social network sites (Hu et al., 2014). A teenage girl was the only person present in the picture. Social comparison requires similarity between the observer and the persons that is observed (Suls, Martin, & Wheeler, 2002), implying that that the girls participating in the study are more likely to compare themselves with females having a comparable age. In Dutch society, the majority of the population is native Dutch and predominantly has a light skin color and Caucasian ethnicity (Centraal Bureau voor de Statistiek, 2015). We, therefore, only used sample images from women with a light skin color. The selected photos may therefore increase the change of social comparison, which is important in light of the study aim. Inspired by Fardouly et al. (2015), another criterion that was applied to the selection of stimulus materials was that photos varied in the parts of the body that were emphasized. Five photos particularly emphasized the girl's face, skin, and hair; the other five highlighted the whole body (see examples in Figure 1).

To create the manipulated photos, each original photo was edited. To this end, effects and filters that are available on Instagram were used. Instagram provides a high number of options to improve pictures. Possibilities include, but are not limited to, improving the color intensity, brightness, and adding strong shadow. Moreover, according to frequently used altering techniques (Philly Renfrew Center Foundation, 2014), we edited the faces and bodies visible in the photos, by removing eye bags, wrinkles, and impurities, and by reshaping legs to be thinner and waist to be slimmer. Finally, all photos were displayed in the same Instagram format. However, we removed comments

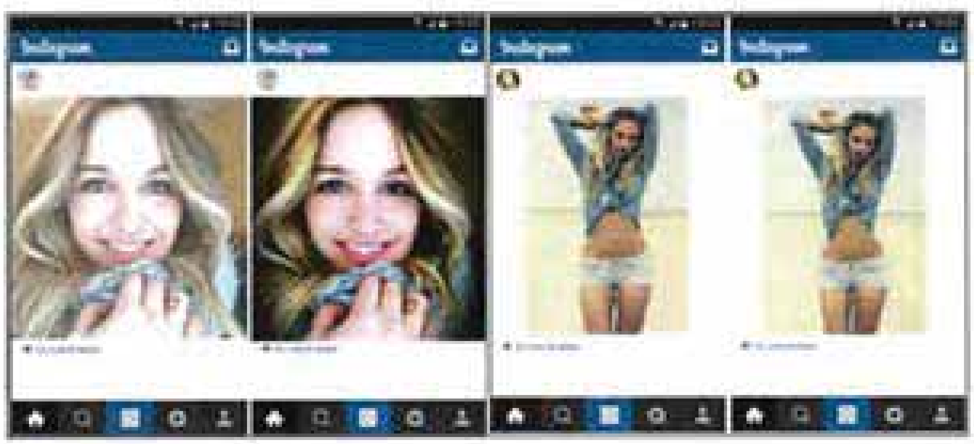


Figure 1. Examples of original versus manipulated Instagram photos emphasizing face, skin, and hair (left), or body (right).

that are normally presented along with photos on Instagram, and gave all materials the same number of likes to exclude this factor as a possible confounder.

Measures

Body image was the dependent variable in the study. The Body Image State Scale (Cash, Fleming, Alindogan, Steadman, & Whitehead, 2002) was used to measure girls' evaluation and affect about their physical appearance at this moment. Girls indicated their (dis)satisfaction with their overall physical appearance; (dis)satisfaction with their body size and shape; (dis) satisfaction with their own weight; feelings of physical (un)attractiveness; current feelings about their own looks relative to how one usually feels; and their evaluation of their appearance relative to how the average person looks. Following Cash et al. (2002), a 9-point, bipolar, Likert-type scale was used with a higher score indicating a more positive body image. As expected, results of a factor analysis including the six items yielded one factor. The initial eigenvalue of this factor (3.341) indicated that this factor explained 55.69% of the variance (factor loadings >.47). In addition, Cronbach's alpha was sufficient (α = .83). We, therefore, calculated the participant's mean score on the statements to construct the variable body image (M = 4.68; SD = 1.26).

To measure girls' *social comparison tendency*, the Iowa-Netherlands Comparison Orientation Measure (Gibbons & Buunk, 1999) was used. This scale consists of 11 items, measured with a 5-point Likert scale ranging from (1) *totally disagree* to (5) *totally agree*. These 11 items were:

(1) I often compare myself with others with respect to what I have accomplished in life;
(2) If I want to learn more about something, I try to find out what others think about it;
(3) I always pay a lot of attention to how I do things compared with how others do things;
(4) I often compare how my loved ones (boy or girlfriend, family members, etc.) are doing with how others are doing;
(5) I always like to know what others in a similar situation would do;
(6) I am not the type of person who compares often with others;
(7) If I want to find out how well I have done something, I compare what I have done with how others have done;
(8) I often try to find out what others think who face similar problems as I face;
(9) I often like to talk with others about mutual opinions and experiences;

(10) I never consider my situation in life relative to that of other people; and

(11) I often compare how I am doing socially (e.g., social skills, popularity) with other people.

The scores on item 6 and item 10 were reversed prior to the analyses. Results of a factor analysis yielded two dimensions. However, the second dimension consisted of only one statement: I often like to talk with others about mutual opinions and experiences. We decided to exclude this item and, thus, to construct the variable based on the remaining ten items (α = .87). Additionally, we created two groups (lower vs. higher tendency) by using a mean split (M = 3.22; SD = .90) to make this variable suitable for the analysis. As a result, 63 girls were indicated as having a lower tendency to make social comparisons, the other 81 girls as having a higher tendency to compare themselves with others.

Level of education is included as control variable in the analysis, as a correlation was found between educational level and the dependent variable (r = .388; p < .001). Those with a higher level of education generally had a more positive body image.

Results

Manipulation checks

For the manipulation check, we asked the 144 girls that participated in the study to respond on a scale ranging from (1) *totally disagree* to (5) *totally agree* to several statements about the photos. First, we asked them to what extend they agreed with the statement that the Instagram photos were manipulated by using filters. Results of a t test showed that their agreement with this statement was higher for the manipulated photos (M = 4.51; SD = .77) than for the original photos (M = 2.19; SD = 1.21), t(142) = −13.759; p < .001. In addition, girls gave higher agreement to the statement that effects (e.g., adding color to look less pale, improving brightness) were used for the manipulated (M = 4.44; SD = .82) than for the original photos (M = 2.11; SD = 1.15), t(142) = −14.055; p < .001. This implies that we were successful in making a distinction between the original and the manipulated photos in this regard. We also asked the participants whether the faces and bodies in the photos were manipulated in terms of reshaping. t tests showed that it was harder to detect these adaptations in the manipulated photos, as the differences compared to the original photos were only marginally significant. For faces, manipulated photos scored somewhat higher (M = 1.82; SD = .81) than the original photos (M = 1.61; SD = .70), t(142) = −1.674; p = .051. In addition, participants slightly more agreed with the statement that bodies were reshaped for the manipulated (M = 1.76; SD = .83) than for the original M = 1.60; SD = .64) photos, t(142) = −1.347; p = .090.

Descriptive Results

Prior to reporting the results of the hypotheses testing, we provide some general information about the girls' evaluation of the photos. First, results of a *t* test showed that girls in the manipulated photos condition rated the photos as more pretty on a 5-point Likert scale ($M = 4.25$; $SD = .69$) than girls in the original photos condition ($M = 3.75$; $SD = .62$), $t(142) = -4.577$; $p < .001$. In addition, the manipulated Instagram photos were perceived as more attractive ($M = 4.57$; $SD = 1.69$) than the original photos ($M = 3.38$; $SD = .67$), $t(142) = -7.533$; $p < .001$. We also found that girls are generally unaware that Instagram photos might be manipulated. To be more specific, for both original and manipulated photos, they agree with the statement that the photos provide a representative view of reality ($M_{original} = 3.68$; $SD_{original} = 1.11$ vs. $M_{manipulated} = 3.72$; $SD_{manipulated} = 1.20$ on a 5-point Likert scale ranging from (1) *totally disagree* to (5) *totally agree*), $t(142) = -.216$; $p = .829$. In addition, no differences were found regarding the statement that the photos paint a picture that is better than reality, $t(142) = -.718$; $p = .474$. The means for both original ($M = 2.22$; $SD = .109$) and manipulated photos ($M = 2.36$; $SD = 1.23$) showed that they generally disagree with this statement.

Effects of manipulated instagram photos

To test the hypotheses, a one-way analysis of covariance was performed with body image as dependent variable, Instagram photo manipulation and tendency to make social comparisons as between-subjects factors, and level of education as covariate. Participant age was excluded as additional covariate because preliminary analyses showed no effects of age on the dependent variable body image. Hypotheses were tested at the alpha = .05 level (one-tailed).

The first hypothesis predicted that girls would have lower body satisfaction after exposure to manipulated Instagram photos than original photos. This hypothesis was supported, $F(1,139) = 4.252$; $p = .021$; $r = .17$. Girls exposed to the manipulated photos showed to have a significant lower body satisfaction ($M = 4.57$; $SE = .13$) compared to girls exposed to the original photos ($M = 4.94$; $SE = .13$). Results additionally showed that level of education (included as control variable) significantly affected body image, $F(1,139) = 14.618$; $p < .001$; $r = .31$. Descriptive statistics showed that a higher the level of education correlates with a more positive body image.

The second hypothesis concerned the moderating effect of the tendency to make social comparisons. First, a main effect of social comparison tendency on body image was found, $F(1,139) = 18.828$; $p < .001$; $r = .35$. Girls who have a higher tendency to compare themselves with others have a lower body

image (M = 4.35; SE = .12) compared to those who have a lower social comparison tendency (M = 5.15; SE = .14). In addition, results provided support for the expectation that the negative effect of manipulated Instagram photos on body image exposure are stronger for girls with a higher social comparison tendency, $F(1,139) = 3.890$; $p = .025$; $r = .16$.

As shown in Figure 2, post-hoc F tests (pair-wise comparisons with Bonferroni correction of the interaction categories) indicate that the body image of girls with a higher tendency to make social comparisons is more negatively affected by manipulated Instagram photos (M = 3.98; SE = .17) than by original photos (M = 4.72; SE = .18), $F(1,139) = 9.209$; $p = .002$; r = . 25. In contrast, girls with a lower tendency to make social comparisons did not significantly differ in body image after exposure to either manipulated (M = 5.16; SE = .18) or original photos (M = 5.15; SE = .21), $F(1,139) = .001$; $p = .485$; $r = .00$. Moreover, results revealed that the negative effect of manipulated photos is more prevalent among girls with a higher tendency to make social comparisons, $F(1,139) = 18.777$; $p < .001$; $r = .34$. Original photos also affected the body image of these girls more compared to ones with a lower tendency to make social comparisons, but this influence is weaker, $F(1,139) = 3.016$; $p = .043$; $r = .15$. In all, the effect of manipulated Instagram

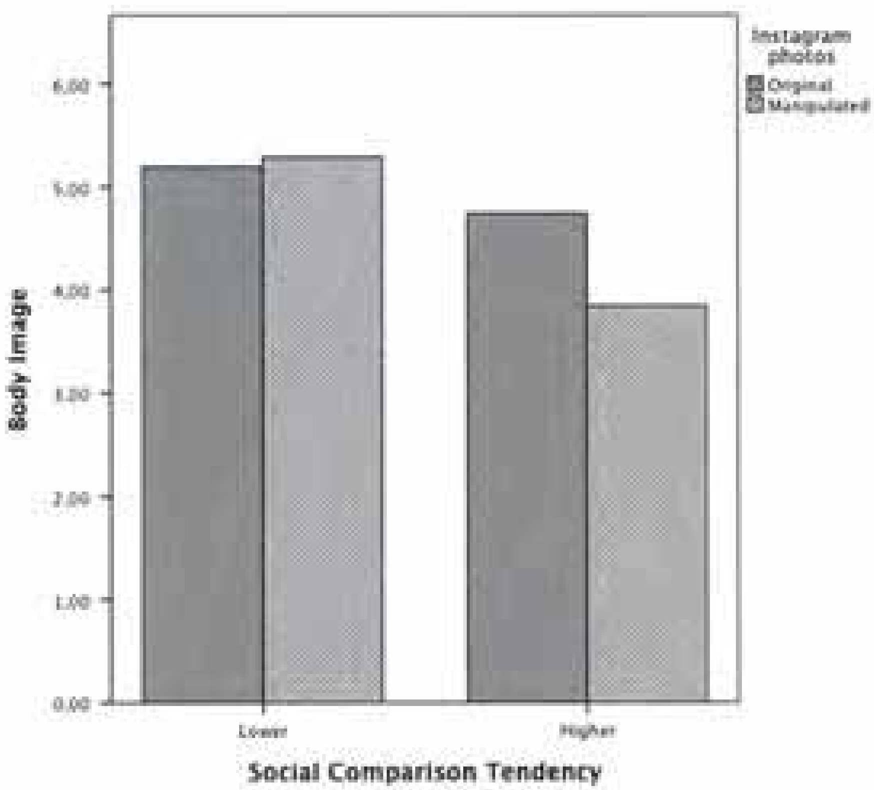


Figure 2. Effect of manipulated versus original Instagram photos on body image among girls with a lower and higher social comparison tendency.

photos on body image exposure is stronger for girls with higher social comparison tendencies.

Discussion

The current study set out to investigate whether manipulated Instagram photos have a negative effect on the body image of female adolescents and whether those with a higher tendency towards social comparison are more vulnerable in this regard. It can be concluded from the results that exposure to manipulated Instagram photos indeed leads to lower body satisfaction in comparison to exposure to non-manipulated selfies from online peers. This particularly related to girls with a higher tendency to make social comparisons. The body image of girls with a lower tendency to compare themselves with others was about equal after exposure to either original or manipulated Instagram photos. In contrast, girls with a higher tendency to make social comparisons had a lower body image in general, and especially after exposure to the manipulated Instagram photos.

These results imply that the common practice of Instagram users to manipulate and tweak their appearance in pictures can have negative consequences, at least for the girls who are prone to make social comparisons. It is worrisome that even short exposure to unfamiliar peers in a research setting can lead to direct changes in body image. The fact that girls believed that the presented Instagram photos showed a representative view of reality and did not notice reshaping of the bodies very well reinforces these concerns. Adolescence is a critical period for psychosocial development and earlier research showed that girls in this phase are more vulnerable for media influences because they equate their own bodies with media images (e.g., Borzekowski et al., 2000). The frequent use of social media networks such as Instagram among young girls (Seetharaman, 2015) clearly stresses the importance of studying the effects of exposure. These results might imply recommending including a disclosure when opening an Instagram account that would remind users that the images on Instagram are often retouched and manipulated, as a means of visual literacy and thereby possible protection from harmful effects. However, following the results of the study by Harrison and Hefner (2014), who reported harmful effects (i.e. lower physical self-esteem and higher body consciousness) of these so called retouched-aware photos, this recommendation might lead to undesirable effects. Therefore, more research is needed to unravel how to best protect these young girls from the negative effects of retouched (social) media images.

The findings of the current study are in line with results found in studies on the effects of exposure to idealized thin bodies in traditional media as well as the first studies on social media networks (i.e. Fardouly et al., 2015; Fardouly & Vartanian, 2015). These studies showed that exposure to

idealized media images lead to a greater focus on the body and more uncertainty among females (Hargreaves & Tiggeman, 2004, Thompson & Heinberg, 1999), which consequently may lead to higher body dissatisfaction (Knauss, Paxton, & Alsaker, 2008). An important difference, however, is that the girls in the current study compared themselves with similar young women (unfamiliar peers), and not with models or celebrities showing the well-known unrealistic beauty standards. In line with earlier findings, the effects of social comparisons may be stronger when perceived similarity is high, which might be the case with exposure to images of peers in social media (see also Andsager et al., 2006; Montoya, Horton, & Kirchner, 2008). Fardouly and colleagues (2015) also argue that the appearance of peers in social media environments is seen as more attainable and, therefore, and more directly triggers social comparison. A suggestion for future research would be to include measures of perceived similarity and attainability to examine this assumption.

Limitations of the current study may also help to shape the future research agenda. First, the current study investigated a short-term effect of exposure to Instagram photos. Therefore, it remains unclear whether this effect will also exist in the longer term. As girls and young women are frequently exposed to Instagram photos (cf. Instagram, 2015), the effect might be even stronger in the long run. Future research should provide more insight into the long-term effects of exposure to Instagram photos on the body image of girls. This research should also strive to assess the general social media habits of the girls involved, to validate if market research positioning of Instagram as the preferred type of social media for girls this age is accurate (Piper Jaffray, 2014; Turpijn, Kneefel, & van der Veer, 2015). In addition, it would be interesting to focus on the effect of Instagram photos presenting people that participants know personally, instead of (only) exposing them to people they are not familiar with. Girls using Instagram are also frequently exposed to photographs of people from their own network, such as friends, classmates, and peers (Madden et al., 2013). One might argue that it is more likely that girls realize that photos are manipulated when they are exposed to photos of people they know, as it is easier to notice that the person in Instagram photos look different than in reality. As a consequence, it is possible that the manipulated photos have a less negative effect on the body image of girls. However, it is also possible that girls have a higher tendency to compare themselves with people from their personal network compared to people they do not know, because of higher perceived similarity and social relevance.

Related to this, our results showed that girls were—at least to some extent—unable to truly detect the retouching of the bodies of the pictures. Although this is in accordance with "real" media pictures (e.g., in magazines) in which retouching is used, this might still be a limitation of our

study because we cannot totally rule out the possibility that our manipulation was too subtle to detect and the effects are solely explained by other factors (e.g., the filters used). However, we considered the manipulation substantial enough to be a valid reflection of reality and the fact that we found differences in body satisfaction shows that although the manipulation might not be truly explicitly perceived, it still can affect the viewer. Following up on this issue, an interesting discussion in this light is how reality is exactly defined. Edited and retouched photos might have become so widely accepted and, therefore, normal for contemporary teenagers (e.g., Choi, 2016; Sutton, Brind & McKenzie, 2007; Wheeler, 2005), that it is hard to tell whether these pictures actually deviate from their view of reality as an issue separate from whether the retouching is detected or not. Related to the recent development that not only celebrities and models, but also peers and teens themselves can idealize their images through retouching, it might even be the case that the distinction between these groups (celebrities vs. peers) in terms of identification and comparison becomes much less profound as the differences between the images become smaller. To investigate this hypothesis, future studies could focus on further examining the changing roles of celebrities versus peers as targets of (appearance-related) comparisons.

Furthermore, the participants were told that the study goal was to investigate how contextual factors affect preferences for different face types, and that they, therefore, would be exposed to pictures of people with different facial expressions. Although we believe that our cover story distracted the participants from (guessing) the real purpose of the study, we are aware of the fact that demand still might have played a role, as exposure to the pictures was followed by the social comparison and body image questions. Demand has been studied in this line of research and some evidence suggests that participants tend to engage in upward comparisons when the study purpose is more obvious rather than less (Mills et al., 2002). It is possible that upward comparisons were stimulated in the present study by asking questions about making comparisons after exposure, although demand characteristics were not explicitly present like in the experiment by Mills et al. (2002). Still, it is very important to pay attention to this topic and to try to avoid demand characteristics as much as possible in future studies. It would also be helpful to ask the participants at the end of the study what they consider the goal of the study, to investigate to what extent they are aware of the purpose. In addition, future studies should include a control condition in which participants are exposed to neutral pictures (e.g., showing landscapes) or not exposed to pictures at all (cf. Harrison & Hefner, 2014), to establish baseline scores on body image in the sample. These baseline scores can serve as a true reference point, since exposure to selfies (edited or not) in itself can affect body image.

Another suggestion for future research would be to include a measure that specifically investigates appearance-related comparison tendencies, for example the Physical Appearance Comparison Scale-Revised (PACS-R, Schaefer & Thompson, 2014). In the current study, we focused on general social comparison tendencies, but we know from the literature that social comparison can manifest itself in many different target domains (Wood, 1989) of which appearance is one that might be particularly interesting in the light of the present study. Using an appearance-related comparison measure might result in even stronger moderating effects than general comparison tendencies, as scoring high on this specific domain of social comparison suggests that girls might have a higher risk of being influenced by social media photos picturing ideal bodies as they specifically compare themselves with others with respect to appearances (e.g., Myers & Crowther, 2009). A final recommendation for future research would be to include ethnicity of the participants as a possible moderator and also include a larger variety of ethnicities in the stimulus material. It is important to include ethnicity in future research because research has shown that both weight-related and general appearance body image varies among ethnic groups (cf. Altabe, 1998; Cachelin, Rebeck, Chung, & Pelayo, 2002; Miller et al., 2000).

In sum, this study contributes to the existing literature regarding media influence on body image in adolescent girls by examining the effects of exposure to Instagram photos of peers. Photo and video sharing becomes more and more common among (young) social network users, so it is important to establish its effects. In addition, the results of the present study add to the public discussion about the use of retouching and reshaping techniques in social media self-presentation materials. The findings indicate that not only celebrities exert influence because they serve as role models, but we should also seriously consider the influence of (unfamiliar) peers.

References

Altabe, M. (1998). Ethnicity and body image: Quantitative and qualitative analysis. *International journal of eating disorders, 23*(2), 153–159. doi:10.1002/(SICI)1098-108X(199803)23:2<153::AID-EAT5>3.0.CO;2-J

Andsager, J. L., Bemker, V., Choi, H., & Torwel, V. (2006). Perceived similarity of exemplar traits and behavior: Effects on message evaluation. *Communication Research, 33*, 3–18. doi:10.1177/0093650205283099

Austin, E. W., & Meili, H. K. (1994). Effects of interpretations on televised alcohol portrayals on children's alcohol beliefs. *Journal of Broadcasting & Electronic Media, 38*, 417–435. doi:10.1080/08838159409364276

Borzekowski, D. L. G., Robinson, T. N., & Killen, J. D. (2000). Does the camera add 10 pounds? Media use, perceived importance of appearance and weight concerns among teenage girls. *Journal of Adolescent Health, 26*, 36–41. doi:10.1016/S1054-139X(99)00044-0

Bandura, A. (2002). Social cognitive theory of mass communication. In J. Bryant & D. Zillmann, *Media effects: Advances in theory and research* (pp. 121–153). Mahwah, NJ: Erlbaum.

Cachelin, F. M., Rebeck, R. M., Chung, G. H., & Pelayo, E. (2002). Does ethnicity influence body-size preference? A comparison of body image and body size. *Obesity Research*, *10*(3), 158–166. doi:10.1038/oby.2002.25

Cash, T., Fleming, E. C., Alindogan, J., Steadman, L., & Whitehead, A. (2002). Beyond body image as a trait: the development and validation of the body image states scale. *Eating Disorders*, *10*, 103–113. doi:10.1080/10640260290081678

Centraal Bureau voor de Statistiek. (2015). *Statline*. Retrieved from http://statline.cbs.nl/Statweb/

Choi, M. H. K. (2016, August). *Like. Flirt. Ghost: A journey into the social media lives of teens*. Retrieved from http://www.wired.com/2016/08/how-teens-use-social-media/

Dittmar, H., & Howard, S. (2004). Thin-ideal internalization and social comparison tendency as moderators of media models' impact on women's body-focused anxiety. *Journal of Social and Clinical Psychology*, *23*,768–791. doi:10.1521/jscp23.6.768.54799

Dreisbach, S. (2014). *How do you feel about your body?* Retrieved from http://www.glamour.com/health-fitness/2014/10/body-image-how-do-you-feel-about-your-body

Fardouly, J., Diedrichs, P. C., Vartanian, L. R., & Halliwell, E. (2015). Social comparisons on social media: The impact of Facebook on young women's body image concerns and mood. *Body Image*, *13*, 38–45. doi:10.1016/j.bodyim.2014.12.002

Fardouly, J., & Vartanian, L. R. (2015). Negative comparisons about one's appearance mediate the relationship between Facebook usage and body image concerns. *Body Image*, *12*, 82–88. doi:10.1016/j.bodyim.2014.10.004

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*, 117–140. doi:10.1177/001872675400700202

Gibbons, F. X., & Buunk, B. P. (1999). Individual differences in social comparison: Development of a scale of social comparison orientation. *Journal of Personality and Social Psychology*, *76*(1), 129–142. doi:10.1037/0022-3514.76.1.129

Grabe, S., Ward, L. M., & Hyde, J. S. (2008). The role of the media in body image concerns among women: a meta-analysis of experimental and correlational studies. *Psychological Bulletin*, *134*(3), 460–476. http://dx.doi.org/10.1037/0033-2909.134.3.460

Halliwell, E., & Dittmar, H. (2004). Does size matter? The impact of model's body size on women's body-focused anxiety and advertising effectiveness. *Journal of Social and Clinical Psychology*, *23*(1), 104–122. doi:10.1521/jscp.23.1.104.26989

Hargreaves, D. A., & Tiggemann, M. (2004). Idealized media images and adolescent body image: "comparing" boys and girls. *Body image*, *1*(4), 351–361. doi:10.1016/j.bodyim.2004.10.002d

Harrison, K., & Hefner, V. (2014). Virtually perfect: Image retouching and adolescent body image. *Media Psychology*, *17*(2), 134–153. doi:10.1080/15213269.2013.770354

Henderson-King, E., & Henderson-King, D. (1997). Media effects on women's body esteem: Social and individual difference factors. *Journal of Applied Social Psychology*, *27*, 399–417. doi: 10.1111/j.1559-1816.1997.tb00638.x

Hu, Y., Manikonda, L., & Kambhampati, S. (2014). What we Instagram: A first analysis of Instagram photo content and user types. In Association for the Advancement of Artificial Intelligence (Ed.), *Eighth International AAAI Conference on Weblogs and Social Media* (pp. 595–598). Palo Alto, CA: The AAAI Press. Retrieved from http://rakaposhi.eas.asu.edu/instagram-icwsm.pdf

Instagram (2015). *Instagram statistics*. Retrieved from http://instagram.com/press

Irving, L. M. (1990). Mirror images: Effects of the standard of beauty on the self-and body-esteem of women exhibiting varying levels of bulimic symptoms. *Journal of Social and Clinical Psychology*, 9(2), 230–242. doi:10.1521/jscp.1990.9.2.230

Jones, D.C. (2001). Social comparison and body image: Attractiveness comparisons to models and peers among adolescent girls and boys. *Sex Roles*, 45, 645–664. doi:10.1023/A:1014815725852

Joshi, R., Herman, C. P., & Polivy, J. (2004). Self-enhancing effects of exposure to thin-body images. *International Journal of Eating Disorders*, 35, 333–341. doi:10.1002/eat.10253

Keery, H., van den Berg, P., & Thompson, J. K. (2004). An evaluation of the tripartite model of body dissatisfaction and eating disturbance with adolescent girls. *Body Image*, 1, 237–251. doi:10.1016/j.bodyim.2004.03.001

Knauss, C., Paxton, S. J., & Alsaker, F. D. (2008). Body dissatisfaction in adolescent boys and girls: Objectified body consciousness, internalization of the media body ideal and perceived pressure from media. *Sex Roles*, 59(9–10), 633–643. doi:10.1007/s11199-008-9474-7

Mabe, A. G., Forney, K. J. and Keel, P. K. (2014). Do you "like" my photo? Facebook use maintains eating disorder risk. *International Journal of Eating Disorders*, 47, 516–523. doi:10.1002/eat.22254

Madden, M., Lenhart, A., Cortesi, S., Gasser, U., Duggan, M., Smith, A., & Beaton, M. (2013). *Teens, social media and privacy*. Washington, DC: Pew Internet and American Life Project.

Manago, A. M., Graham, M. B., Greenfield, P. M., & Salimkhan, G. (2008). Self presentation and gender on MySpace. *Journal of Applied Developmental Psychology*, 29(6), 446–458. doi:10.1016/j.appdev.2008.07.001

Miller, K. J., Gleaves, D. H., Hirsch, T. G., Green, B. A., Snow, A. C., & Corbett, C. C. (2000). Comparisons of body image dimensions by race/ethnicity and gender in a university population. *International Journal of Eating Disorders*, 27(3), 310–316. doi:10.1002/(SICI)1098-108X

Miller, D. T., Turnbull, W., & McFarland, C. (1988). Particularistic and universalistic evaluation in the social comparison process. *Journal of Personality and Social Psychology*, 55, 908–917. doi:10.1037/0022-3514.55.6.908

Mills, J. S., Polivy, J., Herman, C. P., & Tiggemann, M. (2002). Effects of exposure to thin media images: Evidence of self-enhancement among restrained eaters. *Personality and Social Psychology Bulletin*, 28, 1687–1699. doi:10.1177/014616702237650

Montoya, R. M., Horton, R. S., & Kirchner, J. (2008). Is actual similarity necessary for attraction? A meta-analysis of actual and perceived similarity. *Journal of Social and Personal Relationships*, 25, 889–922. doi:10.1177/0265407508096700

Myers, P. N., & Biocca, F. A. (1992). The elastic body image: The effects of television advertising and programming on body image distortions in young women. *Journal of Communication*, 42, 108–133. doi:10.1111/j.1460-2466.1992.tb00802.x

Myers, T. A., & Crowther, J. H. (2009). Social comparison as a predictor of body dissatisfaction: A meta-analytic review. *Journal of Abnormal Psychology*, 18, 683–698. http://dx.doi.org/10.1037/a0016763

Philly Renfrew Center Foundation. (2014). *Afraid to be your selfie? Survey reveals most people Photoshop their images*. Retrieved from http://renfrewcenter.com/news/afraid-be-your-selfie-survey-reveals-most-people-photoshop-their-images

Piper Jaffray. (2014). *Taking stock with teens—2014*. Retrieved from http://www.piperjaffray.com/2col.aspx?id=178

Rainie, L., Brenner, J., & Purcell, K. (2012). *Photos and videos as social currency online*. Pew Internet & American Life Project. Retrieved from http://www.pewinternet.org/2012/09/13/photos-and-videos-as-social-currency-online/

Saltz, J. (2014,January 26). Art at arm's length: A history of the selfie. *New York Magazine*. Retrieved from http://www.vulture.com/2014/01/history-of-the-selfie.html

Sass, E. (2014, October). Social media fueling women's body image issues. *Mediapost*. Retrieved from http://www.mediapost.com/publications/article/236998/social-media-fueling-womens-body-image-issues.html

Schaefer, L. M., & Thompson, J. K. (2014). The development and validation of the physical appearance comparison scale-revised (PACS-R). *Eating Behaviors*, *15*(2), 209–217. doi:10.1016/j.eatbeh.2014.01.001

Scheerens, J., Luyten, H., & Van Ravens, J. (2011). Description and earlier quality review of the Dutch educational system (primary and secondary education). In J. Scheerens H. Luyten, & J. Van Ravens (Eds.) *Perspectives on educational quality* (pp. 53–69). Dordrecht, Germany: Springer.

Seetharaman, D. (2015, October). *Survey finds teens prefer Instagram, Twitter, Snapchat for social networks*. Retrieved from http://blogs.wsj.com/digits/2015/10/16/survey-finds-teens-prefer-instagram-snapchat-among-social-networks/

Strahan, E. J., Wilson, A. E., Cressman, K. E., & Buote, V. M. (2006). Comparing to perfection: How cultural norms for appearance affect social comparisons and self-image. *Body Image*, *3*(3), 211–227. doi:10.1016/j.bodyim.2006.07.004

Sturdevant, M., & Spear, B. (2002). Adolescent psychosocial development. *Journal of the American Dietetic Association*, *102*, 30–31. doi:10.1016/S0002-8223(02)90419-

Sullivan, R. (2014). *Celebrities are using photo manipulation apps to make themselves look thinner*. Retrieved from http://www.news.com.au/lifestyle/beauty/celebrities-are-using-photo-manipulation-apps-to-make-themselves-look-thinner/story-fnjcnzwg-1226874550847

Suls, J., Martin, R., & Wheeler, L. (2002). Social comparison: Why, with whom, and with what effect. *Current Directions in Psychological Science*, *11*, 159–163. doi:10.1111/1467-8721.00191

Sutton, D., Brind, S., & McKenzie, R. (2007). *The state of the real: aesthetics in the digital age*. London, UK: IB Tauris.

Thompson, J. K., & Heinberg, L. J. (1999). The media's influence on body image disturbance and eating disorders: We've reviled them, now can we rehabilitate them? *Journal of Social Issues*, *55*(2), 339–353.

Thornton, B., & Maurice, J. (1999). Physical attractiveness contrast effect and the moderating influence of self-consciousness. *Sex Roles*, *40*, 379–392. doi:10.1023/A:1018867409265

Thornton, B., & Moore, S. (1993). Physical attractiveness contrast effect: Implications for self-esteem and evaluations of the social self. *Personality and Social Psychology Bulletin*, *19*, 474–480. doi:10.1177/0146167293194012

Tiggemann, M., & Miller, J. (2010). The Internet and adolescent girls' weight satisfaction and drive for thinness. *Sex Roles*, *63*, 79–90. doi:10.1007/s11199-010-9789-z

Tiggemann, M., & Slater, A. (2013). NetGirls: The Internet, Facebook, and body image concern in adolescent girls. *International Journal of Eating Disorders*, *46*, 630–634. doi:10.1002/eat.22141

Turpijn, L., Kneefel, S., & van der Veer, N. (2015). *Nationale social media onderzoek 2015* [National Social Media Research 2015]. Amsterdam, The Netherlands: Newcom Research & Consultancy.

Vartanian, L. R., & Dey, S. (2013). Self-concept clarity, thin-ideal internalization, and appearance-related social comparison as predictors of body dissatisfaction. *Body Image*, 10, 495–500. http://dx.doi.org/10.1016/j.bodyim.2013.05.004

Wheeler, T. H. (2005). *Phototruth or photofiction?: Ethics and media imagery in the digital age*. New York, NY: Routledge.

Wilcox, K., & Laird, J. D. (2000). The impact of media images of super-slender women on women's self-esteem: Identification, social comparison, and self-perception. *Journal of Research in Personality*, *34*, 278–288. doi:10.1006/jrpe.1999.2281

Winter, J. (2013, July). Selfie-loathing. Instagram is even more depressing than Facebook. Here's why. *Slate*. Retrieved from http://www.slate.com/articles/technology/technology/2013/07/instagram_and_self_esteem_why_the_photo_sharing_network_is_even_more_depressing.html

Won Kim, J., & Chock T. M. (2015). Body image 2.0: Associations between social grooming on Facebook and body image concerns. *Computers in Human Behavior*, *48*, 331–339. doi:10.1016/j.chb.2015.01.009

Wood, J. V. (1989). Theory and research concerning social comparisons of personal attributes. *Psychological Bulletin*, *106*, 231–248.

Yamamiya, Y., Cash, T. F., Melnyk, S. E., Posavac, H. D., & Posavac, S. S. (2005). Women's exposure to thin- and beautiful media images: Body image effects of media ideal internalization and impact-reduction interventions. *Body Image*, *2* (1), 74–80. doi:10.1016/j.bodyim.2004.11.001

Zillmann, D., & Brosius, H.B. (2000). *Exemplification in communication: The influence of case reports on the perception of issues*. Mahwah, NJ: Erlbaum.

9

American Economic Review 2020, 110(3): 629–676
https://doi.org/10.1257/aer.20190658

The Welfare Effects of Social Media[†]

By Hunt Allcott, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow*

The rise of social media has provoked both optimism about potential societal benefits and concern about harms such as addiction, depression, and political polarization. In a randomized experiment, we find that deactivating Facebook for the four weeks before the [illegible] *and online activity, while increas-* [illegible] *and socializing with* [illegible] *news knowledge and* [illegible] *subjective well-being; and* [illegible] *experiment Facebook* [illegible] *valuations of Facebook,* [illegible] *ate consumer surplus.*

Experimental

S[illegible] modern world. Facebook, which ren[illegible] has 2.3 billion monthly active us[illegible] average user was spending 50 mi[illegible] rms Instagram and Messenger (F[illegible] e television that has so dramati- ca[illegible] spend their time.

Specula[illegible] t has followed a familiar trajectory, with early optimism about pot[illegible] giving way to widespread concern about possible harms. At a basic level, social media dramatically reduce the cost of connecting, communicating, and sharing information with others. Given that interpersonal connections are among the most important drivers of happiness and

*Allcott: New York University and NBER (email: hunt.allcott@nyu.edu); Braghieri: Stanford University (email: lucabrag@stanford.edu); Eichmeyer: Stanford University (email: saraeich@stanford.edu); Gentzkow: Stanford University and NBER (email: gentzkow@stanford.edu). Esther Duflo was the coeditor for this article. We thank Nancy Baym, Moira Burke, Annie Franco, Alex Leavitt, Todd Rogers, Joel Waldfogel, and seminar participants at Berkeley, Centro de Investigación y Docencia Económicas, Chicago, Columbia, Instituto Tecnológico Autónomo de México, Microsoft, NBER Digitization, NYU, Stanford, the Technology Policy Institute, UCLA, UC Santa Cruz, and University of Washington for helpful comments. We thank Raj Bhargava, Zong Huang, and Kelly B. Wagman for exceptional research assistance. We are grateful to the Sloan Foundation and the Knight Foundation for generous support. The study was approved by Institutional Review Boards at Stanford (eProtocol 45403) and NYU (IRB-FY2018-2139). This RCT was registered in the American Economic Association Registry for randomized control trials under trial number AEARCTR-0003409. Replication files and survey instruments are available from https://sites.google.com/site/allcott/research. Disclosures: Allcott is a paid employee of Microsoft Research. Gentzkow does perform paid consulting work for Amazon and is a member of the Toulouse Network for Information Technology, a research group funded by Microsoft. Braghieri and Eichmeyer have no relevant or material disclosures.
[†]Go to https://doi.org/10.1257/aer.20190658 to visit the article page for additional materials and author disclosure statements.

American Economic Review 2020, 110(3): 629–676
https://doi.org/10.1257/aer.20190658

The Welfare Effects of Social Media†

By Hunt Allcott, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow*

The rise of social media has provoked both optimism about potential societal benefits and concern about harms such as addiction, depression, and political polarization. In a randomized experiment, we find that deactivating Facebook for the four weeks before the 2018 US midterm election (i) reduced online activity, while increasing offline activities such as watching TV alone and socializing with family and friends; (ii) reduced both factual news knowledge and political polarization; (iii) increased subjective well-being; and (iv) caused a large persistent reduction in post-experiment Facebook use. Deactivation reduced post-experiment valuations of Facebook, suggesting that traditional metrics may overstate consumer surplus. (*JEL* D12, D72, D90, I31, L82, L86, Z13)

Social media have had profound impacts on the modern world. Facebook, which remains by far the largest social media company, has 2.3 billion monthly active users worldwide (Facebook 2018). As of 2016, the average user was spending 50 minutes per day on Facebook and its sister platforms Instagram and Messenger (Facebook 2016). There may be no technology since television that has so dramatically reshaped the way people get information and spend their time.

Speculation about social media's welfare impact has followed a familiar trajectory, with early optimism about potential benefits giving way to widespread concern about possible harms. At a basic level, social media dramatically reduce the cost of connecting, communicating, and sharing information with others. Given that interpersonal connections are among the most important drivers of happiness and

*Allcott: New York University and NBER (email: hunt.allcott@nyu.edu); Braghieri: Stanford University (email: lucabrag@stanford.edu); Eichmeyer: Stanford University (email: saraeich@stanford.edu); Gentzkow: Stanford University and NBER (email: gentzkow@stanford.edu). Esther Duflo was the coeditor for this article. We thank Nancy Baym, Moira Burke, Annie Franco, Alex Leavitt, Todd Rogers, Joel Waldfogel, and seminar participants at Berkeley, Centro de Investigación y Docencia Económicas, Chicago, Columbia, Instituto Tecnológico Autónomo de México, Microsoft, NBER Digitization, NYU, Stanford, the Technology Policy Institute, UCLA, UC Santa Cruz, and University of Washington for helpful comments. We thank Raj Bhargava, Zong Huang, and Kelly B. Wagman for exceptional research assistance. We are grateful to the Sloan Foundation and the Knight Foundation for generous support. The study was approved by Institutional Review Boards at Stanford (eProtocol 45403) and NYU (IRB-FY2018-2139). This RCT was registered in the American Economic Association Registry for randomized control trials under trial number AEARCTR-0003409. Replication files and survey instruments are available from https://sites.google.com/site/allcott/research. Disclosures: Allcott is a paid employee of Microsoft Research. Gentzkow does perform paid consulting work for Amazon and is a member of the Toulouse Network for Information Technology, a research group funded by Microsoft. Braghieri and Eichmeyer have no relevant or material disclosures.

†Go to https://doi.org/10.1257/aer.20190658 to visit the article page for additional materials and author disclosure statements.

well-being (Myers 2000; Reis, Collins, and Berscheid 2000; Argyle 2001; Chopik 2017), this could be expected to bring widespread improvements to individual welfare. Many have also pointed to wider social benefits, from facilitating protest and resistance in autocratic countries, to encouraging activism and political participation in established democracies (Howard et al. 2011, Kirkpatrick 2011).

More recent discussion has focused on an array of possible negative impacts. At the individual level, many have pointed to negative correlations between intensive social media use and both subjective well-being and mental health.[1] Adverse outcomes such as suicide and depression appear to have risen sharply over the same period that the use of smartphones and social media has expanded.[2] Alter (2018) and Newport (2019), along with other academics and prominent Silicon Valley executives in the "time well-spent" movement, argue that digital media devices and social media apps are harmful and addictive. At the broader social level, concern has focused particularly on a range of negative political externalities. Social media may create ideological "echo chambers" among like-minded friend groups, thereby increasing political polarization (Sunstein 2001, 2017; Settle 2018). Furthermore, social media are the primary channel through which misinformation spreads online (Allcott and Gentzkow 2017), and there is concern that coordinated disinformation campaigns can affect elections in the United States and abroad.

In this paper, we report on a large-scale randomized evaluation of the welfare impacts of Facebook, focusing on US users in the run-up to the November 2018 midterm elections. We recruited a sample of 2,743 users through Facebook display ads, and elicited their willingness-to-accept (WTA) to deactivate their Facebook accounts for a period of four weeks ending just after the election. We then randomly assigned the 61 percent of these subjects with WTA less than $102 to either a Treatment group that was paid to deactivate, or a Control group that was not. We verified compliance with deactivation by regularly checking participants' public profile pages. We measured a suite of outcomes using text messages, surveys, emails, direct measurement of Facebook and Twitter activity, and administrative voting records. Less than 2 percent of the sample failed to complete the endline survey, and the Treatment group's compliance with deactivation exceeded 90 percent.

Our study offers the largest-scale experimental evidence available to date on the way Facebook affects a range of individual and social welfare measures. We evaluate the extent to which time on Facebook substitutes for alternative online and offline activities, with particular attention to crowd out of news consumption and face-to-face social interactions. We study Facebook's broader political externalities via measures of news knowledge, awareness of misinformation, political engagement, and political polarization. We study the impact on individual utility via measures of subjective well-being, captured through both surveys and text messages. Finally, we analyze the extent to which forces like addiction, learning, and projection bias may cause suboptimal consumption choices, by looking at how usage and valuation of Facebook change after the experiment.

[1] See, for example, Vanden Abeele et al. (2018); Burke and Kraut (2016); Ellison, Steinfield, and Lampe (2007); Frison and Eggermont (2015); Kross et al. (2013); Satici and Uysal (2015); Shakya and Christakis (2017); and Tandoc, Ferrucci, and Duffy (2015). See Appel, Gerlach, and Crusius (2016) and Baker and Algorta (2016) for reviews.

[2] See, for example, Twenge, Sherman, and Lyubomirsky (2016); Twenge and Park (2019); Twenge, Martin, and Campbell (2018); and Twenge et al. (2018).

Our first set of results focuses on substitution patterns. A key mechanism for effects on individual well-being would be if social media use crowds out face-to-face social interactions and thus deepens loneliness and depression (Twenge 2017). A key mechanism for political externalities would be if social media crowds out consumption of higher-quality news and information sources. We find evidence consistent with the first of these but not the second. Deactivating Facebook freed up 60 minutes per day for the average person in our Treatment group. The Treatment group actually spent less time on both non-Facebook social media and other online activities, while devoting more time to a range of offline activities such as watching television alone and spending time with friends and family. The Treatment group did not change its consumption of any other online or offline news sources and reported spending 15 percent less time consuming news.

Our second set of results focuses on political externalities, proxied by news knowledge, political engagement, and political polarization. Consistent with the reported reduction in news consumption, we find that Facebook deactivation significantly reduced news knowledge and attention to politics. The Treatment group was less likely to say they follow news about politics or the President, and less able to correctly answer factual questions about recent news events. Our overall index of news knowledge fell by 0.19 standard deviations. There is no detectable effect on political engagement, as measured by voter turnout in the midterm election and the likelihood of clicking on email links to support political causes. Deactivation significantly reduced polarization of views on policy issues and a measure of exposure to polarizing news. Deactivation did not statistically significantly reduce affective polarization (i.e., negative feelings about the other political party) or polarization in factual beliefs about current events, although the coefficient estimates also point in that direction. Our overall index of political polarization fell by 0.16 standard deviations. As a point of comparison, prior work has found that a different index of political polarization rose by 0.38 standard deviations between 1996 and 2018 (Boxell 2018).

Our third set of results looks at subjective well-being. Deactivation caused small but significant improvements in well-being, and in particular in self-reported happiness, life satisfaction, depression, and anxiety. Effects on subjective well-being as measured by responses to brief daily text messages are positive but not significant. Our overall index of subjective well-being improved by 0.09 standard deviations. As a point of comparison, this is about 25–40 percent of the effect of psychological interventions including self-help therapy, group training, and individual therapy, as reported in a meta-analysis by Bolier et al. (2013). These results are consistent with prior studies suggesting that Facebook may have adverse effects on mental health. However, we also show that the magnitudes of our causal effects are far smaller than those we would have estimated using the correlational approach of much prior literature. We find little evidence to support the hypothesis suggested by prior work that Facebook might be more beneficial for "active" users: for example, users who regularly comment on pictures and posts from friends and family instead of just scrolling through their news feeds.[3]

[3]Correlation studies on active versus passive Facebook use include Burke, Marlow, and Lento (2010); Burke, Kraut, and Marlow (2011); Burke and Kraut (2014); and Krasnova et al. (2013), and randomized experiments include Deters and Mehl (2013) and Verduyn et al. (2015).

Our fourth set of results considers whether deactivation affected people's demand for Facebook after the study was over, as well as their opinions about Facebook's role in society. As the experiment ended, participants reported planning to use Facebook much less in the future. Several weeks later, the Treatment group's reported usage of the Facebook mobile app was about 11 minutes (22 percent) lower than in Control. The Treatment group was more likely to click on a post-experiment email providing information about tools to limit social media usage, and 5 percent of the Treatment group still had their accounts deactivated nine weeks after the experiment ended. Our overall index of post-experiment Facebook use is 0.61 standard deviations lower in Treatment than in Control. In response to open-answer questions several weeks after the experiment ended, the Treatment group was more likely to report that they were using Facebook less, had uninstalled the Facebook app from their phones, and were using the platform more judiciously. Reduced post-experiment use aligns with our finding that deactivation improved subjective well-being, and it is also consistent with the hypotheses that Facebook is habit forming in the sense of Becker and Murphy (1988) or that people learned that they enjoy life without Facebook more than they had anticipated.

Deactivation caused people to appreciate Facebook's both positive and negative impacts on their lives. Consistent with our results on news knowledge, the Treatment group was more likely to agree that Facebook helps people to follow the news. About 80 percent of the Treatment group agreed that deactivation was good for them, but they were also more likely to think that people would miss Facebook if they used it less. In free response questions, the Treatment group wrote more text about how Facebook has both positive and negative impacts on their lives. The opposing effects on these specific metrics cancel out, so our overall index of opinions about Facebook is unaffected.

Our work also speaks to an adjacent set of questions around how to measure the economic gains from free online services such as search and media.[4] In standard models with consumers who correctly optimize their allocation of time and money, researchers can approximate the consumer surplus from these services by measuring time use or monetary valuations, as in Brynjolfsson and Oh (2012); Brynjolfsson, Eggers, and Gannamaneni (2018); Corrigan et al. (2018); and others. But if users do not understand the ways in which social media could be addictive or make them unhappy, these standard approaches could overstate consumer surplus gains. Sagioglu and Greitemeyer (2014) provides suggestive evidence: while their participants predicted that spending 20 minutes on Facebook would make them feel better, it actually caused them to feel worse. Organizations such as Time to Log Off argue that a 30-day "digital detox" would help people align their social media usage with their own best interest.

To quantify the possibility that deactivation might help the Treatment group to understand ways in which their use had made them unhappy, we elicited willingness-to-accept at three separate points, using incentive-compatible Becker-DeGroot-Marschak (1964) mechanisms. First, on October 11, we elicited WTA to deactivate Facebook for weeks 1–4 of the experiment, between October 12

[4] See, for example, Brynjolfsson and Saunders (2009); Byrne, Fernald, and Reinsdorf (2016); Nakamura, Samuels, and Soloveichik (2016); Brynjolfsson, Rock, and Syverson (2019); and Syverson (2017).

and November 8. We immediately told participants the amount that they had been offered to deactivate ($102 for the Treatment group, $0 for Control), and thus whether they were expected to deactivate over that period. We then immediately elicited WTA to deactivate Facebook for the next four weeks *after* November 8, i.e., weeks 5–8. When November 8 arrived, we then re-elicited WTA to deactivate for weeks 5–8. The Treatment group's change in valuation for weeks 5–8 reflects a time effect plus the effect of deactivating Facebook. The Control group's parallel valuation change reflects only a time effect. Thus, the difference between how Treatment versus Control change their WTAs for deactivation for weeks 5–8 reflects projection bias, learning, or other unanticipated experience effects from deactivation.[5]

After weighting our sample to match the average US Facebook user on observables, the median and mean willingness-to-accept to deactivate Facebook for weeks 1–4 were $100 and $180, respectively. These valuations are larger than most estimates in related work by Brynjolfsson, Eggers, and Gannamaneni (2018); Corrigan et al. (2018); Mosquera et al. (2018); and Sunstein (forthcoming). A standard consumer surplus calculation would aggregate the mean valuation across the estimated 172 million US Facebook users, giving $31 billion in consumer surplus from four weeks of Facebook. However, consistent with our other results that deactivation reduced demand for Facebook, deactivation caused WTA for weeks 5–8 to drop by up to 14 percent. This suggests that traditional consumer surplus metrics overstate the true welfare gains from social media, though a calculation that adjusts for the downward WTA revision would still imply that Facebook generates enormous flows of consumer surplus.

What do our results imply about the overall net welfare impact of Facebook? On the one hand, Facebook deactivation increased subjective well-being, and 80 percent of the Treatment group reported that deactivation was good for them. On the other hand, participants were unwilling to give up Facebook unless offered fairly large amounts of money: even after they had deactivated for four weeks, which should have allowed at least some learning or "detox" from addiction. It is not entirely clear whether one should prioritize the survey measures or monetary valuations as normative measures of consumer welfare. Benjamin et al. (2012) suggests that subjective well-being measures like ours are not a complete measure of what people are trying to maximize when they make decisions, but Bohm, Lindén, and Sonnegård (1997); Mazar, Kőszegi, and Ariely (2014); and other studies make clear that monetary valuations are not closely held and can be easily manipulated. We think of these tensions as fodder for future research.

Our results should be interpreted with caution, for several reasons. First, effects could differ with the duration, time period, or scale of deactivation. A longer period without Facebook might have less impact on news knowledge as people find alternative news sources, and either more or less impact on subjective well-being. Effects might be different for our pre-election deactivation than for deactivation in other periods. Furthermore, the effects of deactivating a large share of Facebook users

[5] This measurement connects to the literature on habit formation and projection bias, including Acland and Levy (2015); Becker and Murphy (1988); Becker, Grossman, and Murphy (1991); Busse et al. (2015); Charness and Gneezy (2009); Conlin, O'Donoghue, and Vogelsang (2007); Fujiwara, Meng, and Vogl (2016); Gruber and Kőszegi (2001); Hussam et al. (2016); Loewenstein, O'Donoghue, and Rabin (2003); and Simonsohn (2010).

would likely be different due to network effects, so our parameters are most relevant for individuals independently determining their own Facebook use. Second, our sample is not fully representative. Our participants are relatively young, well-educated, and left-leaning compared to the average Facebook user; we included only people who reported using Facebook more than 15 minutes per day; and people willing to participate in our experiment may also differ in unobservable ways. Third, many of our outcome variables are self-reported, adding scope for both measurement error and experimenter demand effects. However, Section IVF finds no evidence of demand effects, and our non-self-reported outcomes paint a similar picture to the survey responses.

The causal impacts of social media have been of great interest to researchers in economics, psychology, and other fields. We are aware of 12 existing randomized impact evaluations of Facebook.[6] The most closely related is the important paper Mosquera et al. (2018), which was made public the month before ours. That paper also uses Facebook deactivation to study news knowledge and well-being, finding results broadly consistent with those reported here. Online Appendix Table A1 details these experiments in comparison to ours. Our deactivation period is substantially longer and our sample size an order of magnitude larger than most prior experimental work, including Mosquera et al. (2018). We measure impacts on a relatively comprehensive range of outcomes, and we are the only one of these randomized trials to have submitted a pre-analysis plan. Given the effect sizes and residual variance in our sample, we would have been unlikely to have sufficient power to detect any effects if limited to the sample sizes in previous experiments. Our work also relates to quasi-experimental estimates of social media effects by Müller and Schwarz (2018) and Enikolopov, Makarin, and Petrova (2018).

Sections I through III present the experimental design, descriptive statistics, and empirical strategy. Section IV presents the impact evaluation, and Section V discusses measurement of the consumer surplus generated by Facebook.

I. Experimental Design

A. *Experiment Overview*

Figure 1 summarizes our experimental design and time line. We timed the experiment so that the main period of Facebook deactivation would end shortly after the 2018 US midterm elections, which took place on November 6. The experiment has eight parts: recruitment, pre-screen, baseline survey, midline survey, endline survey, post-endline survey, post-endline emails, and daily text messages.

Between September 24 and October 3, we recruited participants using Facebook ads. Our ad said, "Participate in an online research study about internet browsing and

[6] These studies sit within a broader media effects literature that uses experimental and quasi-experimental methods to quantify the effects of media technologies such as television, media providers such as Fox News, and content such as political advertising (Bartels 1993; Besley and Burgess 2001; DellaVigna and Kaplan 2007; Enikolopov, Petrova, and Zhuravskaya 2011; Gentzkow 2006; Gerber and Green 2000; Gerber et al. 2011; Gerber, Karlan, and Bergan 2009; Huber and Arceneaux 2007; Martin and Yurukoglu 2017; Olken 2009; and Spenkuch and Toniatti 2016). For reviews, see DellaVigna and Gentzkow (2010), Napoli (2014), Strömberg (2015), Enikolopov and Petrova (2015), and DellaVigna and La Ferrara (2015).

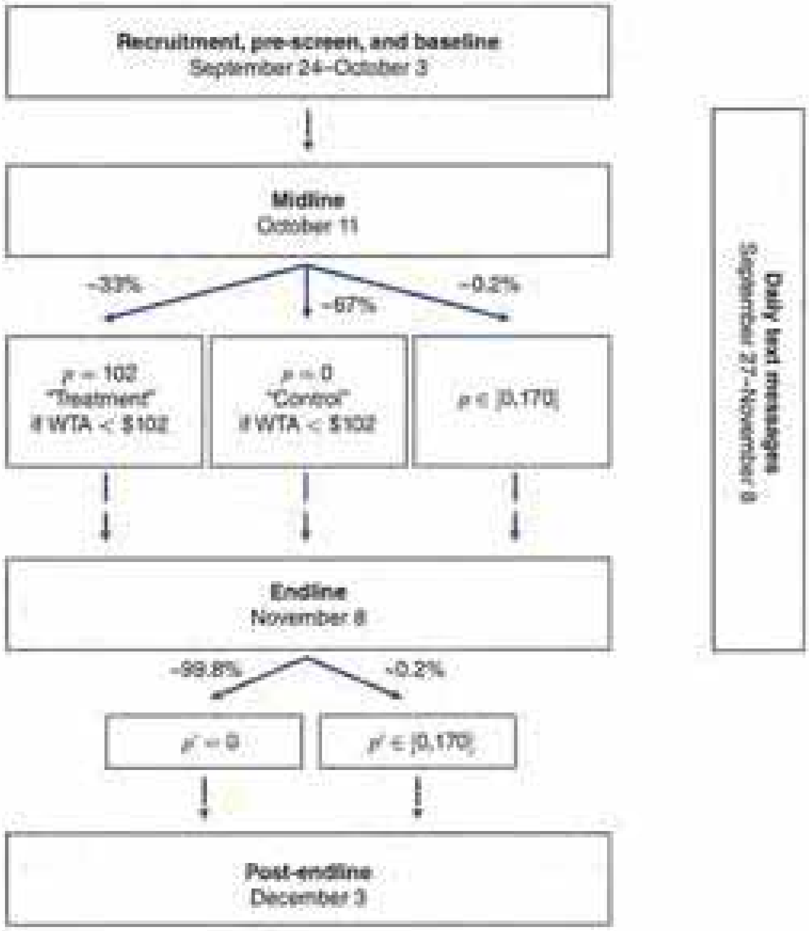


Figure 1. Experimental Design

earn an easy \$30 in electronic gift cards." Online Appendix Figure A1 presents the ad. To minimize sample selection bias, the ad did not hint at our research questions or suggest that the study was related to social media or Facebook deactivation. We targeted the ads by demographic cells in an attempt to gather an initial sample that was approximately representative of Facebook users on gender, age, college completion, and political ideology. A total of 1,892,191 unique users were shown the ad, of whom 32,201 clicked on it. This 1.7 percent click-through rate is about twice the average click-through rate on Facebook ads across all industries.[7]

Clicking on the ad took the participant to a brief pre-screen survey, which included several background demographic questions and the consent form. A total of 17,335 people passed the pre-screen, by reporting being a US resident born between the years 1900 and 2000 who uses Facebook more than 15 minutes and no more than 600 minutes per day. Of those people, 7,455 consented to participate in the study.

After completing the consent form, participants began the baseline survey. The baseline recorded email addresses, additional demographics, and a range of outcome

[7] Mark Irvine, "Facebook Ad Benchmarks for YOUR Industry," WordStream, August 27, 2019, https://www.wordstream.com/blog/ws/2017/02/28/facebook-advertising-benchmarks.

variables. We also asked for each participant's name, zip code, Twitter handle, and phone number ("in order for us to send you text messages during the study"), as well as the URL of their Facebook profile page (which we would use "solely to observe whether your Facebook account is active"). Finally, we informed people that we would later ask them to deactivate their accounts for two 24-hour periods, and confirmed their willingness to do so. (We required all participants regardless of treatment status to deactivate for these 24-hour periods to minimize selective attrition and to ensure that the valuations described below reflect value of Facebook access, not the fixed cost of the deactivation process.)

In all, 3,910 people finished the baseline survey and were willing to deactivate. Of those, 1,013 were dropped from the experiment because of invalid data (for example, invalid Facebook profile URLs) or low-quality baseline responses (for example, discrepancies between average daily Facebook usage reported in the pre-screen versus baseline survey, completing the survey in less than ten minutes, no text in short-answer boxes, and other patterns suggesting careless responses). The remaining 2,897 participants had valid baseline data, were included in our stratified randomization, and were invited to take the midline survey.

On October 11, we sent an email invitation to the midline survey. The survey first asked participants to deactivate their Facebook accounts for 24 hours and guided them through the process. The survey clearly explained what deactivation entailed and how we would monitor deactivation. Facebook allows users to deactivate and reactivate their accounts at any time. We informed participants that they could continue to use Facebook Messenger while deactivated, and that their profile and friend network would be unchanged when they reactivated. We emphasized that Facebook would automatically reactivate their account if they logged into the Facebook website or app, or if they actively logged into any *other* app using their Facebook login credentials.[8] We informed participants that "We will verify whether or not you deactivated your account by pinging the Facebook URL" that they had provided in the baseline survey.

The midline survey then used a Becker-DeGroot-Marschak (BDM) mechanism to elicit willingness-to-accept (WTA) to stay deactivated for four weeks rather than 24 hours.[9] We then revealed the BDM price offer. An additional 154 participants had dropped out before this point of the midline survey, leaving 2,743 who received their price offer. Participants whose WTA was strictly less than the price draw were informed that they should deactivate for the full four weeks after midline. Finally, the midline survey reminded people that we would again ask them to deactivate for

[8] A user's Facebook account automatically reactivates whenever the user actively logs into any other app using their Facebook login credentials. However, this does not fully preclude people from using other apps for which they had used Facebook to log in. People can continue using other apps if they are already logged in, can set up non-Facebook logins, or can log in with Facebook and then again deactivate their Facebook account.

[9] The survey explained, "The computer has randomly generated an amount of money to offer you to deactivate your Facebook account for the next 4 weeks. Before we tell you what the offer is, we will ask you the smallest offer you would be willing to accept. If the offer the computer generated is above the amount you give, we will ask you to deactivate for 4 weeks and pay you the offered amount if you do. If the offer is below that amount, we will not ask you to deactivate." We then asked several comprehension questions to make sure that participants understood the mechanism. We did not tell participants the distribution or support of the offer prices, both because we did not want to artificially truncate the distribution of elicited WTA and because prior studies have found that providing information on the bounds of the offer price distribution can affect BDM valuations (Bohm, Lindén, and Sonnegård 1997; Mazar, Kőszegi, and Ariely 2014).

24 hours after the endline survey, and used a second BDM mechanism to elicit WTA to stay deactivated for the four weeks after endline instead of just 24 hours. We refer to the four weeks after midline as "weeks 1–4," and the four weeks after endline as "weeks 5–8."

On November 8, two days after the midterm election, we sent an email invitation to the endline survey. The endline survey first measured the same outcome variables as the baseline survey. All questions were identical, with the exception of cases discussed in Section IC, such as using updated news knowledge questions and rephrasing questions about the midterm election to be in the past tense. We then asked all participants to again deactivate their Facebook accounts for the next 24 hours, and again elicited WTA to stay deactivated for the next four weeks (i.e., weeks 5–8) instead of the next 24 hours. Participants were told, "With a 50 percent chance we will require you to abide by the decision you made 4 weeks ago; with 50 percent chance we will ignore the decision you made 4 weeks ago and we will require you to abide by the decision you make today."

We gathered data from two post-endline emails. On November 20, we sent an email with links to information on ways to limit smartphone social media use, and on November 25, we sent an email with links to donate to, volunteer for, or sign petitions related to political causes. Clicks on these emails provide additional non-self-reported measures of interest in reducing social media use and political engagement. Online Appendix Figures A2 and A3 present the two emails.

On December 3, we invited participants to a short post-endline survey in which we asked how many minutes per day they had used the Facebook app on their smartphones in the past seven days. We asked participants with iPhones to report the Facebook app time reported by their phone's Settings app, and we asked other participants to estimate. We also asked several open-answer questions, such as "How has the way you use Facebook changed, if at all, since participating in this study?"

For the approximately six weeks between baseline and endline, we sent daily text message surveys to measure several aspects of subjective well-being in real time rather than retrospectively. We rotated three types of questions, measuring happiness, the primary emotion felt over the past ten minutes, and loneliness. Online Appendix Figure A4 presents the three questions.

We verified deactivation by checking each participant's Facebook profile page URL regularly at random times. While a user can limit how much content other people can see in their profiles, they cannot hide their public profile page, and the public profile URL returns a valid response if and only if their account is active.[10] This is thus our measure of deactivation. For all participants, we verified deactivation approximately once per day for the seven days before midline and all days between endline and the end of January 2019. Between midline and endline, we verified deactivation approximately four times per day for people who were supposed to be

[10]By default, Facebook profile URLs end in a unique number, which is the numeric ID for that person in the Facebook system. Users can update their default URL to be something customized, and they can change their customized URL as often as they want. In the baseline survey, participants reported their profile URLs, which could have been either the default or customized version. Shortly after the baseline survey, we checked if each participant's Facebook profile URL was valid by pinging it and looking in the page source for the string containing the person's numeric ID. If the numeric ID existed, we knew that the URL was valid. After that point, we used participants' numeric IDs to construct their default numeric URLs, which allowed us to correctly measure deactivation even if they changed their customized URL.

deactivated (i.e., the Treatment group) and once every four days for everyone else. During the post-midline and post-endline 24-hour deactivation periods, we generally verified deactivation within about six hours of when each participant completed the survey. If participants were not deactivated when they were supposed to be, our program immediately sent an automated email informing them that they should again deactivate as soon as possible, along with a survey asking them to explain why they were not deactivated.

All participants received \$5 per completed survey, paid via gift card immediately upon completion. All participants were told that they would receive a \$15 "completion payment" if they completed all surveys, responded to 75 percent of text messages, kept their accounts deactivated for the 24 hours after midline and endline, and, if the deactivation offer price was above their reported WTA, kept their accounts deactivated for the full period between midline and endline. The latter requirement (making the completion payment contingent on complying with the BDM's deactivation assignment) makes it a strictly dominant (instead of weakly dominant) strategy to truthfully report valuations in the BDM.[11] These payments were in addition to the \$102 that the Treatment group received in exchange for deactivation.

B. *Randomization*

We used the BDM mechanism described above to randomly assign participants to Facebook deactivation. Figure 1 illustrates the randomization. Participants with valid baseline data were randomized into three groups that determined the BDM offer price p for deactivation in weeks 1–4 (i.e., the weeks between midline and endline): $p = \$102$ (approximately 33 percent of the sample), $p = \$0$ (approximately 67 percent), and p drawn from a uniform distribution on $[\$0, \$170]$ (approximately 0.2 percent).[12] We balanced the $p = \$102$ and $p = \$0$ group assignments within 48 strata defined by age, average daily Facebook use, heavy versus light news use (those who get news from Facebook fairly often or very often versus never, hardly ever, or sometimes), active versus passive Facebook use, and Democrat, Republican, or independent party affiliation.

The effects of Facebook deactivation in weeks 1–4 are identified in the sample of participants who were allocated to $p = \$102$ or $p = \$0$ and were willing to accept less than \$102 to deactivate in weeks 1–4. We call this the "impact evaluation sample." Within the impact evaluation sample, we call $p = \$102$ the "Treatment" group, and $p = \$0$ the "Control" group.

For deactivation in weeks 5–8 (i.e., the four weeks after endline), 0.2 percent of participants were randomly selected to a BDM offer price drawn randomly from $p' \in [0, 170]$, while the remaining 99.8 percent received offer $p' = 0$. We balanced

[11] As discussed above, we did not inform participants of the BDM offer price distribution. Thus, more precisely, truthfully reporting valuations is a strictly dominant strategy only within the support of the offer price distribution that participants expected us to use.

[12] We chose \$102 because our pilot data correctly suggested that there would be a point mass of WTAs at \$100 and that it would maximize statistical power per dollar of cost to set an offer price just high enough to induce those participants to deactivate. We chose \$170 as the top of the uniform distribution because it was the maximum that we could pay participants without requiring tax-related paperwork.

this weeks 5–8 offer price p' between the weeks 1–4 offer price groups, so two participants who were offered $p = \$102$ and four participants who were offered $p = \$0$ were assigned to positive weeks 5–8 offers $p' \in [0, 170]$.

This approach allows us to maintain incentive compatibility in the BDM mechanism, have balance between Treatment and Control groups, and use a straightforward regression to estimate treatment effects of post-midline deactivation.

C. *Outcome Variables*

For the impact evaluation, we consider the outcome variables in the nine families described below. Online Appendix B presents survey question text and descriptive statistics for each outcome variable and moderator, grouped by family. We also construct indices that combine the outcome variables within each family, weighting by the inverse of the covariance between variables at endline, as described in Anderson (2008). In constructing these indices, we orient the variables so that more positive values have the same meaning: for example, more positive means "more polarized" in all cases. Outcomes to be multiplied by -1 are followed by "$\times (-1)$" in online Appendix B.

Substitute Time Uses.—At baseline and endline, we asked participants how many minutes per day they spent on Facebook on the average day in the past four weeks. At baseline, we also asked participants to report how much of their free time on the average day in the past four weeks they spent on various activities, ranging from using social media apps other than Facebook to spending time with friends and family in person. At endline, we asked how much time they spent on the same activities, "relative to what is typical for you." We phrased the questions in this way in order to more precisely detect changes in self-reported time use caused by the deactivation.

Social Interaction.—We have three measures of social interaction. The *friends met in person* variable is the natural log of 1 plus the number of friends seen in person in the last week, as measured by a survey question that asked participants to "list the first names of as many friends you met in person last week that you can think of in 1 minute." *Offline activities* is the number of offline activities (such as going out to dinner, spending time with your kids, etc.) that the person did at least once last week. *Diverse interactions* is an indicator for whether the respondent interacted with someone who voted the opposite way in the last presidential election plus an indicator for whether the respondent interacted with someone from another country in the last week.

Substitute News Sources.—At baseline, we asked participants how often they got news from different sources over the past four weeks, including Facebook, cable TV, print, and radio news, borrowing a standard survey question from the Pew Research Center (2018). At endline, we again asked how often they got news from those same sources, "relative to what is typical for you." For the participants who reported having a Twitter handle, we gathered data on number of tweets in the four weeks before baseline began and in the four weeks between midline and endline. This

allows a non-self-reported measure of one kind of potential substitution away from Facebook.[13]

News Knowledge.—In order to detect broad changes in news exposure, we asked participants how closely they followed politics, how closely they followed news about President Trump, and how many minutes per day they spent watching, reading, or listening to the news (including on social media) over the past four weeks.

In order to measure specific news knowledge, we included a 15-question news knowledge quiz. For each question, we gave a statement from the news in the past four weeks and asked participants to indicate if they thought the statement was true or false, or whether they were unsure. The order of the 15 statements was randomized. Seven of the statements were from news stories covered in the past four weeks in six news websites: *New York Times*, *Wall Street Journal*, Fox News, CNN, MSNBC, and *US News & World Report*, such as "The Trump administration set the maximum number of refugees that can enter the country in 2019 to 30,000." Three of the headlines were false modifications of articles from those same six news websites, such as "President Trump spoke at the funeral of former Arizona Senator John McCain, honoring the late McCain's wish." (In reality, it had been reported that President Trump was not invited to McCain's funeral.) The *news knowledge* variable is the count of true statements rated as true plus the count of false statements rated as false, plus one-half for every statement about which the respondent was "unsure." The final five statements were from fake news stories, rated false by third-party fact-checkers Snopes.com and Factcheck.org, that circulated heavily within a four-week period before the survey. The *fake news knowledge* variable is the count of fake statements correctly rated as "false" plus one-half for every statement about which the respondent was unsure. Online Appendix B presents the full news knowledge quizzes from both baseline and endline.

Political Engagement.—We have two measures of political engagement. First, we measure whether participants voted in the 2018 midterm election, by matching participants on name, birth year, and zip code to a voting database supplied to Stanford by L2, a voting data provider. See online Appendix C for details on the match process. Second, we measure whether participants clicked on any of the links in the post-endline politics email.

Political Polarization.—There are a variety of ways to measure political polarization (see, for example, Gentzkow 2016), and we use both standard and novel measures. First, we included standard "feeling thermometer" questions capturing how "warm or cold" participants felt toward the Democratic and Republican Parties and President Trump over the past four weeks. The *party affective polarization* variable is the respondent's thermometer warmth toward her own party minus her warmth toward the other party. For this and all other polarization variables, we include independents who lean toward a party, and we drop independents who do not lean toward either party.

[13] In our pre-analysis plan, we grouped this *number of tweets* variables in the substitute news sources family, but one might also think of it as a "substitute time use" because Twitter is not only used to read news.

Second, the *Trump affective polarization* variable is the thermometer warmth toward President Trump for Republicans, and -1 times the thermometer warmth toward President Trump for Democrats. Third, we asked respondents to list recent news events that made them angry at the Republican or Democratic Party. *Party anger* is the natural log of 1 plus the length (in characters of text) of her response about the other party minus the natural log of 1 plus the length of her response about her own party. Fourth, we asked people how often they saw news that made them better understand the point of view of the Republican Party, and a parallel question for news about the Democratic Party. *Congenial news exposure* is the respondent's answer about her own political party minus her answer for the other party.

Fifth, we asked opinions about nine current political issues, such as "To what extent do you think that free trade agreements between the US and other countries have been a good thing or a bad thing for the United States?" These nine questions were all adapted from recent Pew Center and Gallup opinion polls. The *issue polarization* variable reflects the extent to which the respondent's issue opinions align with the average opinion in her own party instead of the other party. Sixth, *belief polarization* reflects the extent to which the respondent's beliefs about current news events (from the news knowledge quiz described above) align with the average belief in her own party instead of the other party.[14] Finally, *vote polarization* measures the strength of preferences for the congressional candidate of the respondent's party in the midterm election.[15]

Subjective Well-Being.—There is a vast literature on measuring subjective well-being (see, for example, Kahneman et al. 2006), and we use standard measures from the literature. We modified existing scales in two ways. First, we asked questions in reference to the past four weeks, so as to increase our ability to detect changes as a result of Facebook deactivation. Second, in some cases we chose a subset of questions from standard multi-question scales in order to focus on areas of subjective well-being that might be most affected by Facebook.

The *happiness* variable is the average response to two questions from the Subjective Happiness Scale (Lyubomirsky and Lepper 1999), asking how happy participants were over the past four weeks and how happy they were compared to their peers. *Life satisfaction* is the sum of responses to three questions from the Satisfaction with Life Scale (Diener et al. 1985), such as the level of agreement with

[14] Specifically, for each issue or belief question q, we normalize responses by the standard deviation in the Control group, determine Democrats' and Republicans' average responses μ_q^D and μ_q^R, recenter so that $\mu_q^D + \mu_q^R = 0$, and resign so that $\mu^R > 0$. Define $\tilde{y}_{iq}$ as individual i's normalized, recentered, and re-signed response to question q, multiplied by -1 if i is a Democrat. Thus $\tilde{y}_{iq}$ reflects the strength of individual i's agreement with the average view of her party instead of the other party. For *issue polarization*, further define σ_q as the Control group within-person standard deviation of $\tilde{y}_{iq}$ for question q. This measures how much people's views change between baseline and endline, and allows us to place higher weight on issues about which views are malleable over the deactivation period. For belief polarization, let $\sigma_q = 1$. The issue and belief polarization measures are $Y_i = \sum_q \tilde{y}_{iq}\sigma_q$. Online Appendix Table A15 shows that the *issue polarization* results are nearly identical if we set $\sigma_q = 1$.

[15] Specifically, we asked "In the recent midterm elections, did you vote for the Republican Party's or for the Democratic Party's candidate for Congress in your district? (If you did not vote, please tell us whom you would have voted for.)" We code vote polarization as 0 for "other/don't know." For people who responded that they had (or would have) voted for the Republican or Democratic candidate, we then asked, "How convinced were you about whether to vote for the Republican candidate or the Democratic candidate?" In these cases, we code vote polarization on a scale from -1 (very convinced to vote for the Democratic candidate) to $+1$ (very convinced to vote for the Republican candidate), and then multiply by -1 for Democrats.

the statement, "During the past 4 weeks, I was satisfied with my life." *Loneliness* is the Three-Item Loneliness Scale (Hughes et al. 2004). Finally, *depressed, anxious, absorbed*, and *bored* reflect how much of the time during the past four weeks respondents felt each emotion, using questions from the European Social Survey well-being module (Huppert et al. 2009).

The daily text messages allowed us to measure the aspects of subjective well-being that are most important to record in the moment instead of retrospectively. This approach builds on the Experience Sampling Method of Csikszentmihalyi and Larson (2014) and Stone and Shiffman (1994). The variable *SMS happiness* is the answer to the question, "Overall, how happy do you feel right now on a scale from 1 (not at all happy) to 10 (completely happy)?" The variable *SMS positive emotion* is an indicator variable for whether the participant reports a positive emotion when asked, "What best describes how you felt over the last ten minutes?" with possible responses such as "angry," "worried," "loving/tender," etc. Finally, *SMS not lonely* uses the answer to the question, "How lonely are you feeling right now on a scale from 1 (not at all lonely) to 10 (very lonely)?"

Post-Experiment Facebook Use.—We have four measures of planned and actual post-experiment Facebook use. First, *planned post-study use change* is the extent to which participants plan to use Facebook more or less than they had before they started the study. (This was included only in the endline survey.) Second, *clicked time limit email* is an indicator for whether the respondent clicked any of the links in the post-endline social media time limit email. Third, *speed of reactivation* is −1 times the natural log of 1 plus the number of days that the participant's account remained deactivated between the post-endline 24-hour deactivation period and our most recent measurement on December 17. Fourth, *Facebook mobile app use* is the natural log of 1 plus the number of minutes per day that the participant reported using Facebook on their phone in the post-endline survey.

Opinions about Facebook.—We asked eight questions eliciting people's opinions about Facebook, such as "To what extent do you think Facebook is good or bad for society?" and "To what extent do you think Facebook makes people more or less politically polarized?" Each of these eight responses was on a ten-point scale. In the endline survey only, we also asked *Deactivation bad*: "As part of this study, you were asked to deactivate your Facebook account for [24 hours/4 weeks]. To what extent do you think that deactivating your account was good or bad for you?" Finally, we also included two open answer text boxes in which we asked people to write out the most important positive and negative impacts that Facebook has on their lives. The *positive impacts* and *negative impacts* variables are the natural log of 1 plus the count of characters in the respective text box.

Secondary Outcomes.—We also consider the following two outcomes, which we labeled as "secondary" in our pre-analysis plan. First, we consider the standard generic ballot question. At baseline, we asked "If the elections for US Congress were being held today, would you vote for the Republican Party's candidate or the Democratic Party's candidate for Congress in your district?" To increase precision,

TABLE 1—SAMPLE SIZES

Phase	Sample size
Recruitment and baseline	N = 1,892,191 were shown ads N = 32,201 clicked on ads N = 22,324 completed pre-screen survey N = 20,959 were from United States and born between 1900 and 2000 N = 17,335 had $15 <$ daily Facebook minutes ≤ 600 N = 7,455 consented to participate N = 3,910 finished baseline N = 2,897 had valid baseline and were randomized, of which:
Midline	N = 2,897 began midline N = 2,743 received a price offer, of which: N = 1,661 were in impact evaluation sample
Endline	N = 2,710 began endline N = 2,684 finished endline, of which: N = 1,637 were in impact evaluation sample
Post-endline	N = 2,067 reported Facebook mobile app use, of which: N = 1,219 were in impact evaluation sample

we then asked, "How convinced are you about whether to vote for the Republican or Democratic candidate?" At endline, we asked these questions in past tense, about whom the respondent did vote for in the 2018 midterm (or whom the respondent would have voted for had she voted, to avoid potentially selective non-response). The *voted Republican* variable is the strength of preferences for the Republican candidate. We labeled this outcome as secondary because we expected the estimates to be too imprecise to be of interest.

Second, we asked people to report whether they had voted (at endline) and planned to vote (at baseline) in the 2018 midterm. We labeled this as secondary because it is superseded by the administrative voting data from L2.

We also gathered contributions to political campaigns from the Federal Election Commission database. In our pre-analysis plan, we labeled this as secondary because very few Americans contribute to political campaigns, and we did not expect to be able to detect effects from four weeks of deactivation. Indeed, only one person in the impact evaluation sample donated to a political party between the October 2018 midline survey and July 2019. As a result, we deviate from the pre-analysis plan by dropping this from our analysis.

II. Descriptive Statistics

Table 1 shows sample sizes at each step of our experiment, from the 1.9 million Facebook users who were shown our ads, to the 1,661 subjects in the impact evaluation sample. Table 2 quantifies the representativeness of our sample on observables, by comparing the demographics of our impact evaluation sample to our estimate of the average demographics of adult Facebook users and to the US adult population. Comparing column 1 to columns 2 and 3, we see that our sample is relatively high-income, well-educated, female, young, and Democratic, and uses Facebook

TABLE 2—SAMPLE DEMOGRAPHICS

	Impact evaluation sample (1)	Facebook users (2)	US population (3)
Income under $50,000	0.40	0.41	0.42
College	0.51	0.33	0.29
Male	0.43	0.44	0.49
White	0.68	0.73	0.74
Age under 30	0.52	0.26	0.21
Republican	0.13		0.26
Democrat	0.42		0.20
Facebook minutes	74.52	45.00	

Notes: Column 1 presents average demographics for the impact evaluation sample: participants who were willing to accept less than $102 to deactivate Facebook for the four weeks after midline and were offered p = $102 or p = $0 to do so. Column 2 presents our estimate of average demographics of American adults with a Facebook account. The top five numbers in column 2 are inferred from a Pew Research Center (2018f) survey of social media use by demographic group. The bottom number in column 2 (the average of 45 minutes of Facebook use per day) is approximated on those basis of sources such as Facebook (2016) and Molla and Wagner (2018). Column 3 presents average demographics of American adults. The top five numbers are from the 2017 American Community Survey (US Census Bureau 2017), and the Republican and Democrat shares are from the 2016 American National Election Study (American National Election Studies 2016).

relatively heavily.[16] Online Appendix Table A14 shows that Treatment and Control are balanced on observables.

Table 3 documents very high response rates to the endline and post-endline surveys and subjective well-being text messages. Of the 580 people in the Treatment group, only 7 failed to complete the endline survey. Of the 1,081 people in the Control group, only 17 failed to complete endline. The average participant responded to 92 percent of daily text messages, well above the 75 percent required in order to receive the completion payment.[17] Treatment and Control have statistically equal response rates to the endline survey and subjective well-being text messages. A marginally significantly larger share of the Treatment group responded to the post-endline survey; this is less worrisome because Facebook mobile app use is the only variable from that survey for which we calculate treatment effects, and we show in online Appendix Table A13 that using Lee (2009) bounds to account for attrition does not change the conclusions. Finally, Table 3 also reports the high level of compliance with our deactivation treatment: treatment group participants were deactivated on 90 percent of checks between October 13 (the first day after the 24-hour post-midline deactivation period) and November 7 (the day before endline), against 2 percent for Control.

As described above, if Treatment group members were found to have active accounts, we sent an email informing them of this and asking them to promptly deactivate, along with a survey asking why they were not deactivated. From these

[16] In online Appendix Figures A17, A18, A19, and A20, we find that the two demographic variables that we prespecified as moderators, age and political party, do not appear to systematically moderate treatment effects. Furthermore, Figure 9 provides no systematic evidence that the effects vary for people who use Facebook more versus less heavily before baseline. This suggests that reweighting the sample for representativeness on these observables would not substantively change the estimated effects, although it would increase the standard errors.

[17] Online Appendix Figure A26 shows the text message response rate by day (response rates declined slightly over the course of the experiment) and shows that Treatment and Control response rates are statistically balanced in all days of the deactivation period.

TABLE 3—SURVEY RESPONSE AND TREATMENT COMPLIANCE RATES

Variable	Treatment mean/SD (1)	Control mean/SD (2)	t-test p-value (1) − (2)
Completed endline survey	0.99 (0.11)	0.98 (0.12)	0.54
Share of text messages completed	0.92 (0.20)	0.93 (0.18)	0.45
Completed post-endline survey	0.95 (0.23)	0.92 (0.26)	0.07
Share days deactivated	0.90 (0.29)	0.02 (0.13)	0.00
Observations	580	1,081	

Notes: Columns 1 and 2 present survey response and treatment compliance rates for the Treatment and Control groups in the impact evaluation sample: participants who were willing to accept less than \$102 to deactivate Facebook for the four weeks after midline and were offered $p = \$102$ or $p = \$0$ to do so. Column 3 presents p-values of tests of differences in response rates between the two groups.

surveys, along with email interactions and formal qualitative interviews following our summer 2018 pilot study, we conclude that most Treatment group members who did reactivate fall into one of two groups. The first group consists of a small number of users who changed their mind about participating in the experiment and reactivated intentionally. The second group consists of users who briefly reactivated by accident, for example because they logged in to another app or online service using their Facebook account credentials.

Online Appendix Figure A27 shows the cumulative distribution of the share of time deactivated for the Treatment group, and online Appendix Figure A28 shows the distribution of reasons for deactivation among those for whom this share was less than 1. Together, these figures suggest that the small group of intentional reactivators accounts for the vast majority of Treatment group noncompliance. Given this, combined with the fact that the Control group was also found to be deactivated for a small share of weeks 1–4, we will analyze the experiment as a randomized encouragement design.

III. Empirical Strategy

A. *Pre-Analysis Plan*

We submitted our pre-analysis plan on October 12, as this was the final day before the Treatment and Control groups could have begun to differ. We submitted a slightly updated pre-analysis plan on November 7, the day before endline, with only one substantive change: on the basis of data on reasons for non-compliance described above, we specified that our primary specifications would use IV estimates instead of intent-to-treat estimates. The pre-analysis plan specified three things. First, it specified the outcome variables and families of outcome variables as described above, including which specific variables are included in the index for

each family and which outcomes are "secondary." Versions of Figures 2, 3, 5, 6, 7, and 12 appear as figure shells in the pre-analysis plan, although we changed some variable labels as well as the order in which we present the families of outcome variables for expositional purposes. Second, the pre-analysis plan specified the moderators we use when testing for heterogeneous treatment effects, including which moderators are "secondary." Third, it specified the two regression specifications and the estimation sample as described below.

B. *Empirical Strategy*

To estimate the local average treatment effect (LATE) of Facebook deactivation, define Y_i as some outcome measured at endline, and $\mathbf{Y}_i^b$ as a vector including the baseline value of the outcome and the baseline value of the index that includes the outcome.[18] Define D_i as the percent of deactivation checks between October 13 and November 7 that person i is observed to be deactivated. Define $T_i \in \{1,0\}$ as a Treatment group indicator, and ν_s as the vector of the 48 stratum dummies. We estimate local average treatment effects of deactivation using the following regression:

$$Y_i = \tau D_i + \rho \mathbf{Y}_i^b + \nu_s + \varepsilon_i, \tag{1}$$

instrumenting for D_i with T_i. In equation (1), τ measures the local average treatment effect of deactivation for people induced to deactivate by the promised \$102 payment.[19]

The base sample for all regressions is the "impact evaluation sample"; again, participants who were willing to accept less than \$102 to deactivate in weeks 1–4 (the four weeks after midline) and were offered $p = \$102$ or $p = \$0$ to do so. For the political polarization outcomes, the sample includes only Democrats and Republicans, as well as independents who lean toward one party or the other. Sample sizes sometimes differ across outcomes due to missing data: for example, the post-endline survey has higher non-response than the endline survey, and many participants do not have Twitter accounts.

We use robust standard errors in all regressions.

[18] $\mathbf{Y}_i^b$ excludes the baseline value of the outcome for outcomes such as clicks on post-endline emails that do not have a baseline value. $\mathbf{Y}_i^b$ excludes the baseline index when Y_i is not included in an index. When Y_i is an index, $\mathbf{Y}_i^b$ is simply the baseline value of the index.

[19] Facebook deactivation might have a larger impact for people who use Facebook more. Define H_i as person i's average daily hours of Facebook use reported at baseline, winsorized at 120 minutes. We can also estimate the local average treatment effect of deactivation *per hour of daily Facebook use avoided* using the following regression:

$$Y_i = \tau D_i H_i + \beta H_i + \rho \mathbf{Y}_i^b + \nu_s + \varepsilon_i, \tag{2}$$

analogously instrumenting for $D_i H_i$ with $T_i H_i$.

If effects of deactivation are indeed linear in avoided hours of Facebook use, then equation (2) could provide more statistical power than equation (1). On the other hand, if effects are closer to constant in baseline usage and/or H_i is measured with error, then equation (1) will offer more power. In our pre-analysis plan, we specified that we would make either equation (1) or equation (2) our primary specification, depending on which delivered more power. In reality, the results are very similar. Therefore, we focus on equation (1) because it is simpler. Online Appendix E presents results using equation (2).

IV. Impact Evaluation

This section presents treatment effects of Facebook deactivation. The following subsections present estimates for four groups of outcomes: substitution, news and political outcomes, subjective well-being, and post-experiment Facebook use and opinions. We then present heterogeneous treatment effects. Finally, we provide evidence on experimenter demand effects.

In the body of the paper, we present figures with local average treatment effects and 95 percent confidence intervals from estimates of equation (1), with outcome variables Y_i normalized so that the Control group standard deviation equals 1. Online Appendix Tables A10 and A11 provide numerical regression results for all individual outcome variables in both normalized (standard deviation) units, as in the figures, and unnormalized (original) units. Online Appendix Table A12 provides numerical regression results for all nine summary indices. These tables also provide unadjusted p-values and "sharpened" False Discovery Rate (FDR)-adjusted p-values following the procedure of Benjamini, Krieger, and Yekutieli (2006), as outlined by Anderson (2008). The unadjusted p-values are appropriate for readers with a priori interest in one specific outcome. The FDR-adjusted p-values for the individual outcomes limit the expected proportion of false rejections of null hypotheses across all individual outcomes reported in the paper, while the FDR-adjusted p-values for the indices limit the expected proportion of false rejections of null hypotheses across the nine indices. The sharpened FDR-adjusted p-values are less conservative than the unadjusted p-values for p-values greater than about 0.15, and more conservative for unadjusted p-values less than that.

A. *Substitutes for Facebook*

Figure 2 presents treatment effects on substitutes for Facebook: substitute time uses, social interactions, and substitute news sources. Substitution is of interest for two reasons. First, our treatment entails deactivating Facebook *and* also reallocating that time to other activities. Understanding that reallocation is thus crucial for conceptually understanding the "treatment." Second, this substitution helps to understand mechanisms for key effects. One central mechanism through which Facebook might affect psychological well-being is by crowding out face-to-face interactions. However, it's also possible that when people deactivate, they primarily devote their newly available time to other solitary pursuits. Furthermore, a central mechanism for possible political externalities is that social media use crowds out consumption of higher-quality news. However, it's also possible that when people deactivate, they simply get less news overall instead of substituting to other news sources.

The top group of outcomes in Figure 2 measures self-reported time use. Facebook usage was reported in minutes. For all other activities, the endline survey asked respondents how much time they spent on the activity in the last four weeks relative to what is typical for them, on a five-point scale from "A lot less" to "A lot more." For all time use outcomes, "Same" is the average answer in the Control group.

The first row confirms that the treatment indeed reduced Facebook use as intended. At endline, the Control group reported that they had used Facebook for an average of 59.53 minutes per day over the past four weeks, and the local average

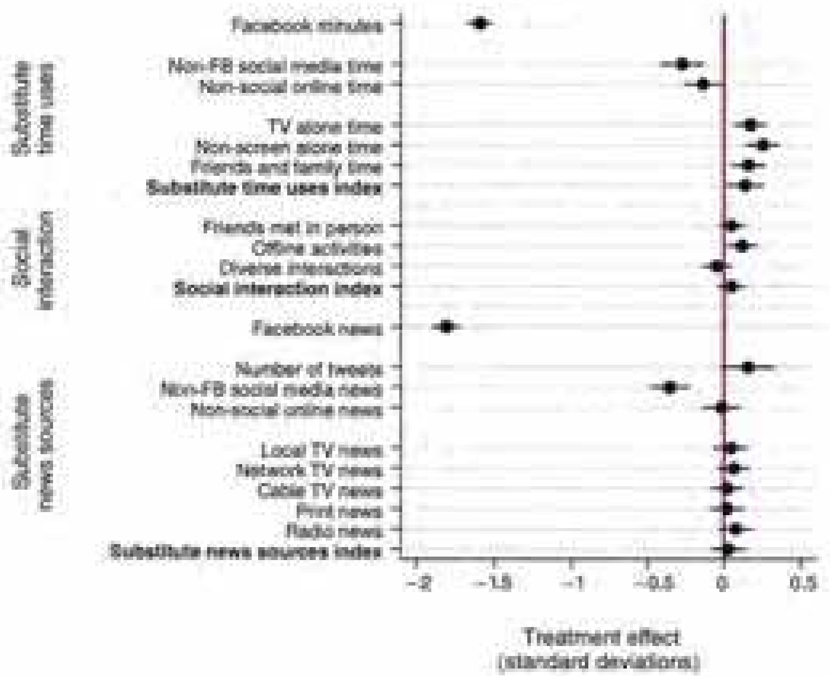


Figure 2. Substitutes for Facebook

Notes: This figure presents local average treatment effects of Facebook deactivation estimated using equation (1). All variables are normalized so that the Control group endline distribution has a standard deviation of 1. Error bars reflect 95 percent confidence intervals. See Section IC for variable definitions. *Facebook minutes* is not included in the substitute time uses index, and *Facebook news* is not included in the substitute news sources index, so we visually separate these two variables from the other variables in their respective families. We also visually separate online and offline time uses and news sources, although all online and offline substitutes enter their respective indexes.

treatment effect of deactivation is 59.58 minutes per day.[20] As shown in Figure 2, this corresponds to a reduction of 1.59 standard deviations.

We find that Facebook deactivation *reduced* time devoted to other online activities. Time using non-Facebook social media falls by a quarter point on our five-point scale (0.27 SD), and time on non-social online activities falls by 0.12 points (0.14 SD). Thus, Facebook appears to be a complement rather than a substitute for other online activities. This makes sense to the extent that deactivating Facebook makes people less likely to be using their phones or computers in the first place, and less likely to follow Facebook links that direct to non-Facebook sites (e.g., a news website or Twitter post). Furthermore, the Treatment group may have avoided logging into other apps such as Spotify and Tinder because we had informed participants that using Facebook to actively log into other apps would reactivate Facebook.

Rows 4–7 of Figure 2 suggest that the 60 minutes freed up by not using Facebook, as well as the additional minutes from reductions in other online activities, were

[20]Online Appendix Table A5 reports baseline means of our time use variables. The mean of self-reported Facebook minutes at baseline is 74.5 minutes per day, and the mean of reported minutes using the Facebook mobile app at baseline is 60 minutes per day.

allocated to both solitary and social activities offline. Solitary television watching increases by 0.17 points on our scale (0.17 SD), other solitary offline activities increase by 0.23 points (0.25 SD), and time devoted to spending time with friends and family increases by 0.14 points (0.16 SD). The substitute time uses index, which does not include *Facebook minutes*, shows an increase in overall non-Facebook activities. All of the online and offline time use effects are highly significant with and without adjustment for multiple hypothesis testing.

The middle group of outcomes in Figure 2 contains measures of social interaction. Deactivation increased the count of offline activities that people reported doing at least once last week by about 0.18 (0.12 SD). Online Appendix Figure A29 shows that the specific activities with the largest point estimates are going out to dinner, getting together with friends, and spending time with parents. The point estimates for the other offline activities we measure (going to the cinema, talking to friends on the phone, going to a party, going shopping, and spending time with your kids) are all very close to zero. Notwithstanding the positive effects on *offline activities*, there are no statistically significant effects on the number of friends that participants listed as having met in person last week, or on *diverse interactions* (whether or not they interacted with someone who voted differently in the last presidential election or interacted with someone from another country). We find no effects on the social interaction index, although the point estimate is positive.

The bottom group of outcomes in Figure 2 measures news consumption. As with the substitute time uses, the endline survey asked participants how much time they spent getting news from each source in the last four weeks relative to what is typical for them; "Same" is again the average answer in the Control group. As expected, Facebook deactivation substantially reduced the extent to which people said they relied on Facebook as a news source. Consistent with the time use results, the Treatment group also got substantially less news from non-Facebook social media sites (0.36 SD). The point estimates for print, radio, and TV news are all positive but statistically insignificant. Facebook deactivation has a positive but insignificant effect on Twitter use. As we discuss below in the news knowledge results, deactivation reduced the total time subjects report spending consuming news by 8 minutes per day, or 15 percent of the Control group mean of 52 minutes.

Overall, these results suggest that Facebook is a substitute for offline activities but a complement to other online activities. This suggests the possibility that Facebook could reduce subjective well-being by reducing in-person interactions, but also impose positive political externalities by increasing news knowledge. Below, we test these possibilities more directly.

B. *Effects on News and Political Outcomes*

Figure 3 presents treatment effects on news and political outcomes: news knowledge, political engagement, and political polarization. News knowledge and political engagement are of interest because well-functioning democratic societies fundamentally rely on well-informed voters who actually show up to the polls to vote. Political polarization is of interest because it is may make democratic decision making less efficient, and may lead citizens to perceive democratic outcomes as less legitimate (Iyengar, Sood, and Lelkes 2012; Iyengar and Westwood 2015).

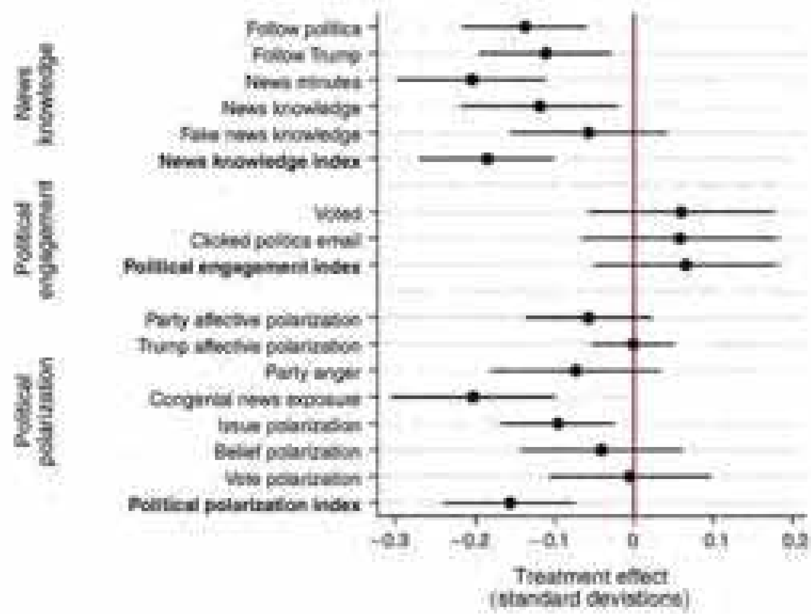


Figure 3. Effects on News and Political Outcomes

Notes: This figure presents local average treatment effects of Facebook deactivation estimated using equation (1). All variables are normalized so that the Control group endline distribution has a standard deviation of 1. Error bars reflect 95 percent confidence intervals. See Section IC for variable definitions.

Deactivation caused substantial reductions in both self-reported attention to news and directly measured news knowledge. The top three rows show that deactivation reduced how much people reported they followed news about politics and about President Trump (by 0.14 and 0.11 SD, respectively), as well as the average minutes per day spent consuming news (a drop of 8 minutes per day, or 15 percent of the control group mean). Accuracy on our news knowledge quiz fell by 0.12 standard deviations.[21] Tangibly, the Control group answered an average of 7.26 out of the 10 news knowledge questions correctly (counting "unsure" as one-half correct), and deactivation reduced this average by 0.14. There is no detectable effect on fake news knowledge, possibly reflecting the limited reach of even the highly shared fake news items included in our survey. Overall, deactivation reduced the news knowledge index by about 0.19 standard deviations.

There are no statistically detectable effects on political engagement. As reported in online Appendix Tables A10 and A11, the point estimates suggest that deactivation increased turnout by three percentage points according to the administrative data and decreased turnout by three percentage points according to the self-reported

[21] Online Appendix G presents more analysis of the effects on news knowledge, including effects on each individual news knowledge and fake news knowledge question. All but one of the point estimates for the ten news knowledge questions is negative. The news knowledge questions with the largest effects involve correctly responding that Elizabeth Warren's DNA test had revealed Native American ancestry and that Jeff Sessions had resigned at President Trump's request. There was also a statistically significant difference in knowledge about one fake news story: the Treatment group was less likely to correctly respond that Cesar Sayoc, the suspect in an act of domestic terrorism directed at critics of President Trump, was not a registered Democrat.

data, and neither estimate is statistically different from zero. Similarly, the Treatment and Control groups are statistically equally likely to have clicked on any link in the post-endline politics email. Online Appendix Figure A35 does show a marginally significant negative effect on *voted Republican*, suggesting that deactivation may have reduced support for Republican congressional candidates. The unadjusted p-value is 0.06, the sharpened FDR-adjusted p-value is 0.08, and we had labeled this as a "secondary outcome" in our pre-analysis plan.

Prior research has shown that people tend to be exposed to ideologically congenial news content in general (Gentzkow and Shapiro 2011) and on Facebook in particular (Bakshy, Messing, and Adamic 2015). Thus, the finding above that deactivation reduced news exposure naturally suggests that deactivation might have also reduced political polarization.

Indeed, deactivation did reduce political polarization. Point estimates are negative for all polarization measures. The largest and most significant individual effect is on *congenial news exposure*: deactivation decreased the number of times that people reportedly saw news that made them better understand the point of view of their own political party relative to the other party. Deactivation also decreased issue polarization, which Fiorina and Abrams (2008) singles out as the "most direct" way of measuring polarization.[22] Online Appendix Table A10 shows that both of these effects are highly significant after adjusting for multiple hypothesis testing. The other measures with the largest point estimates are *party anger* and *party affective polarization*, although these individual effects are not statistically significant. Overall, deactivation reduced the political polarization index by about 0.16 standard deviations.[23]

Figure 4 illustrates how deactivation reduced issue polarization, by plotting the distribution of "issue opinions" for Democrats and Republicans in Treatment and Control at endline. Our *issue opinions* measure exactly parallels the *issue polarization* variable used in the regressions, except that we keep opinions on a left-to-right scale, with more negative indicating more agreement with the average Democratic opinion, and more positive indicating more agreement with the average Republican opinion. (By contrast, the issue polarization variable multiplies Democrats' responses by -1, so that a more positive value reflects more agreement with the average opinion in one's political party.) We then normalize *issue opinions* to have a standard deviation of 1 in the Control group. The figure shows that deactivation moves both Democrats and Republicans visibly toward the center. In the Control group, the issue opinions of the average Democrat and the average Republican differ by 1.47 standard deviations. In the Treatment group, this difference is 1.35 standard deviations: about 8 percent less.

Are these polarization effects large or small? As one benchmark, we can compare these effects to the increase in political polarization in the United States since 1996,

[22] Online Appendix Figure A30 presents results for each of the issue polarization questions. The issues for which deactivation caused the largest decrease in polarization were the direction of racial bias in policing and whether the Mueller investigation is biased.

[23] Like all of our outcome families, the polarization index includes a range of different outcomes with different interpretations. Exposure to congenial news is conceptually different from affective polarization and issue polarization. Online Appendix Table A16 shows that the effect on the political polarization index is robust to excluding each of the seven individual component variables in turn, although the point estimate moves toward zero and the unadjusted p-value rises to 0.09 when omitting *congenial news exposure*.

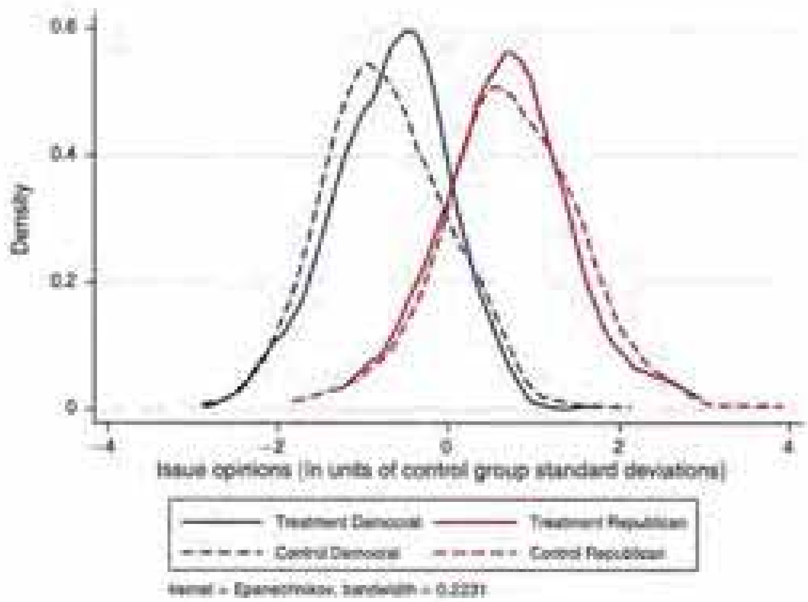


Figure 4. Issue Opinions by Party at Endline

Notes: This figure presents kernel density plots of issue opinions for Democrats and Republicans in Treatment and Control at endline. Issue opinions are attitudes about nine current political issues on a scale from −5 to +5, such as "To what extent do you think that free trade agreements between the US and other countries have been a good thing or a bad thing for the United States." See online Appendix B for a list of all nine issue questions. To construct the *issue opinions* measure, for each issue question q, we normalize responses by the standard deviation in the Control group, determine Democrats' and Republicans' average responses μ_q^D and μ_q^R, recenter so that $\mu_q^D + \mu_q^R = 0$, and resign so that $\mu^R > 0$. Define $\tilde{y}_{iq}$ as individual i's normalized, recentered, and re-signed response to question q. Thus $\tilde{y}_{iq}$ reflects the strength of individual i's agreement with the average Republican. Define σ_q as the Control group within-person standard deviation of $\tilde{y}_{iq}$ for question q. This measures how much people's views change between baseline and endline, and allows us to place higher weight on issues about which views are malleable over the deactivation period. The preliminary *issue opinion* measure is $Y_i = \sum_q \tilde{y}_{iq}\sigma_q$, and the final *issue opinion* measure plotted in the figure is Y_i divided by the Control group standard deviation.

well before the advent of social media. Using data from the American National Election Studies, Boxell (2018) calculates that the change in a different index of polarization measures increased by 0.38 standard deviations between 1996 and 2016. The 0.16 standard deviation effect of Facebook deactivation on political polarization in our sample is about 42 percent as large as this increase.[24]

Overall, these results suggest that Facebook plays a role in helping people stay informed about current events, but also increases polarization, particularly of views on political issues.

[24] Specifically, Boxell's polarization index increased by 0.269 units from 1996–2016, and the standard deviation of Boxell's polarization index across people in 2016 is 0.710 units, so political polarization increased by 0.269/0.71 = 0.379 standard deviations over that period. Of course, this benchmarking exercise does not imply that political polarization in the United States would have increased by one-third less in the absence of Facebook, for many reasons. For example, the treatment effects in our sample from a four-week deactivation are unlikely to generalize to the US population over Facebook's 15-year life. Furthermore, some of our polarization measures are unique to our study. The one measure that appears in both Boxell's index and our index, *party affective polarization*, rose by 0.18 standard deviations between 1996 and 2016. Our point estimate of −0.06 standard deviations is about one-third of this amount, although this estimate is not statistically different from zero.

C. *Effects on Subjective Well-Being*

Figure 5 presents estimates of effects on subjective well-being (SWB). These outcomes are of interest because, as discussed in the introduction, many studies show cross-sectional or time-series correlations between social media use and well-being, and on this basis researchers have speculated that social media may have serious adverse effects on mental health. The outcomes are re-signed so that more positive represents better SWB: for example, the "depressed" variable is multiplied by (−1).

We find that deactivation indeed significantly increases SWB. All but one of the ten point estimates are positive. The magnitudes are relatively small overall, with the largest and most significant effects on *life satisfaction* (0.12 SD), *anxiety* (0.10 SD), *depression* (0.09 SD), and *happiness* (0.08 SD).[29] All of these effects remain significant after adjusting for multiple hypothesis testing. The text message based measures of happiness are not significantly different from zero, with positive point estimates ranging from 0.01 SD to 0.06 SD. Deactivation improved our overall SWB index by 0.09 standard deviations.

Are these subjective well-being effects large or small? As one benchmark, we can consider the effect sizes in their original units, focusing on the measures with the largest effects. *Happiness* is the average response to two questions (for example, "Over the last 4 weeks, I think I was …") on a scale from 1 (not a very happy person) to 7 (a very happy person). The Control group endline average is 4.47 out of a possible 7, and deactivation caused an average increase of 0.12. *Life satisfaction* is the extent of agreement with three questions (for example, "During the past four weeks, I was satisfied with my life") on seven-point Likert scales from "strongly disagree" to "strongly agree." The Control group endline average is 12.26 out of a possible 21, and deactivation caused an average increase of 0.56. *Depressed* and *anxious* are responses to the question, "Please tell us how much of the time during the past four weeks you felt [depressed/anxious]," where 1 is "None or almost none of the time" and 4 is "All or almost all of the time." The average responses are 2.99 and 2.60, respectively, and deactivation caused average increases of 0.08 and 0.09.

As a second benchmark, a meta-analysis of 39 randomized evaluations finds that positive psychology interventions (i.e., self-help therapy, group training, and individual therapy) improve subjective well-being (excluding depression) by 0.34 standard deviations and reduce depression by 0.23 standard deviations (Bolier et al. 2013). Thus, deactivating Facebook increased our subjective well-being index by about 25–40 percent as much as standard psychological interventions.

As a third benchmark, online Appendix Table A17 presents a regression of our baseline SWB index on key demographics (income, college completion, gender, race, age, and political party). College completion is conditionally associated with 0.23 standard deviations higher SWB. Thus, the effect of deactivating Facebook is just over one-third of the conditional difference in subjective well-being between college graduates and everyone else. The table also shows that

[29] Online Appendix Figure A34 presents results for the individual questions within the *happiness*, *life satisfaction*, and *loneliness* scales.

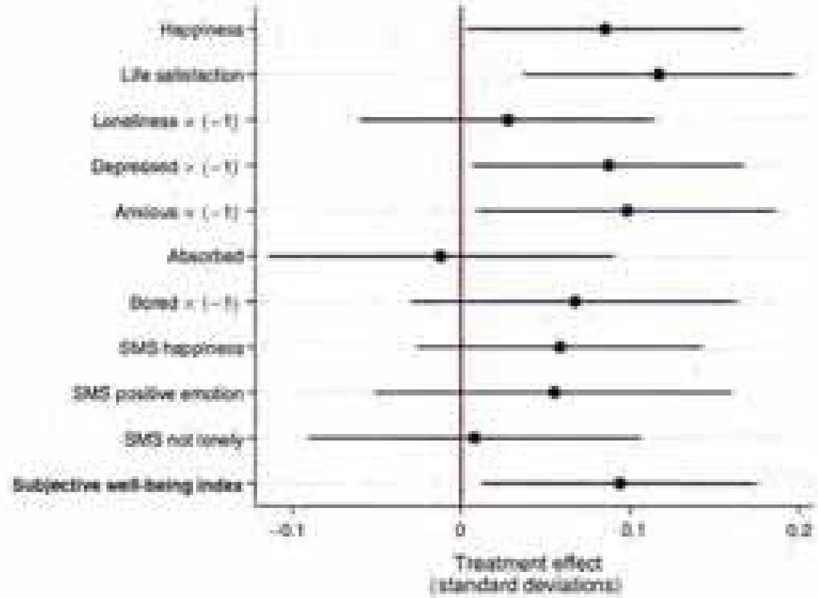


Figure 5. Effects on Subjective Well-Being

Notes: This figure presents local average treatment effects of Facebook deactivation estimated using equation (1). All variables are normalized so that the Control group endline distribution has a standard deviation of 1. Error bars reflect 95 percent confidence intervals. See Section IC for variable definitions.

a $10,000 increase in income is conditionally associated with a 0.027 standard deviation increase in SWB. Thus, the effect of deactivating Facebook is equal to the conditional difference in subjective well-being from about $30,000 additional income. This income equivalent is large because "money doesn't buy happiness": although income is correlated with SWB, the slope of that relationship is not very steep.

Online Appendix Figure A31 presents effects on the SMS outcomes by week of the experiment, to test whether the effects might have some trend over time. None of the effects on any of the three outcomes is statistically significant in any of the four weeks. The point estimates do not systematically increase or decrease over time, and if anything, the point estimates are largest in the first week. This suggests that the effects of a longer deactivation might not be different.

We can also compare our SWB effects to what we would have estimated using the kind of correlational approach taken by many previous non-experimental studies. These studies often have specific designs and outcomes that don't map closely to our paper, so it is difficult to directly compare effect sizes with other papers. We can, however, replicate the empirical strategy of simple correlation studies in our data, and compare our cross-sectional correlations to the experimental results. To do this, we regress SWB outcomes at baseline on daily average Facebook use over the past four weeks as of baseline, divided by the local average treatment effect of deactivation on daily average Facebook use between midline and

endline, so that the coefficients are both in units of average use per day over the past four weeks.[26]

The baseline correlation between our SWB index and Facebook use is about three times larger than the experimental estimate of the treatment effect of deactivation (about 0.23 SD compared to 0.09 SD), and the point estimates are highly statistically significantly different. Controlling for basic demographics brings down the non-experimental estimate somewhat, but it remains economically and statistically larger than our experimental estimate. Online Appendix Figure A32 presents the full results for all SWB outcomes.[27] These findings are consistent with reverse causality, for example if people who are lonely or depressed spending more time on Facebook, or with omitted variables, for example if lower socioeconomic status is associated with both heavy use and lower well-being. They could also reflect a difference between the relatively short-term effects measured in our experiment and the longer-term effects picked up in the cross section. However, the lack of a detectable trend in treatment effects on the text message outcomes over the course of our experiment (as noted above and seen in online Appendix Figure A31) points away from this hypothesis.

Subjects' own descriptions in follow-up interviews and free-response questions are consistent with these quantitative findings, while also highlighting substantial heterogeneity in the effects. Many participants described deactivation as an unambiguously positive experience. One said in an interview,

> *I was way less stressed. I wasn't attached to my phone as much as I was before. And I found I didn't really care so much about things that were happening [online] because I was more focused on my own life ... I felt more content. I think I was in a better mood generally. I thought I would miss seeing everyone's day-to-day activities ... I really didn't miss it at all.*

A second wrote, "I realized how much time I was wasting. I now have time for other things. I've been reading books and playing the piano, which I used to do daily until the phone took over."

A third wrote, "I realized I was using it too much and it wasn't making me happy. I hate all of the interactions I had with people in comment sections."

Many others highlighted ways in which deactivation was difficult. One said in an interview,

> *I was shut off from those [online] conversations, or just from being an observer of what people are doing or thinking ... I didn't like it at first at all, I felt very cut off from people that I like ... I didn't like it because I spend a lot of time by myself anyway, I'm kind of an introvert, so I use Facebook in a social aspect in a very big way.*

[26] Specifically, the non-experimental estimates are from the following regression:

$$Y_i^b = \tau \tilde{H}_i + \beta \mathbf{X}_i + \epsilon_i, \tag{3}$$

where Y_i^b is participant i's value of some outcome measured in the baseline survey, $\mathbf{X}_i$ is a vector of basic demographic variables (household income, age, and college, male, white, Republican, and Democrat indicators), and $\tilde{H}_i$ is baseline average daily Facebook use over the past four weeks (winsorized at 120 minutes per day) divided by the local average treatment effect on average daily Facebook use between midline and endline.

[27] One could also do similar experimental versus non-experimental comparisons for other outcomes, but we have done this only for SWB because SWB is the focus of the non-experimental literature in this area.

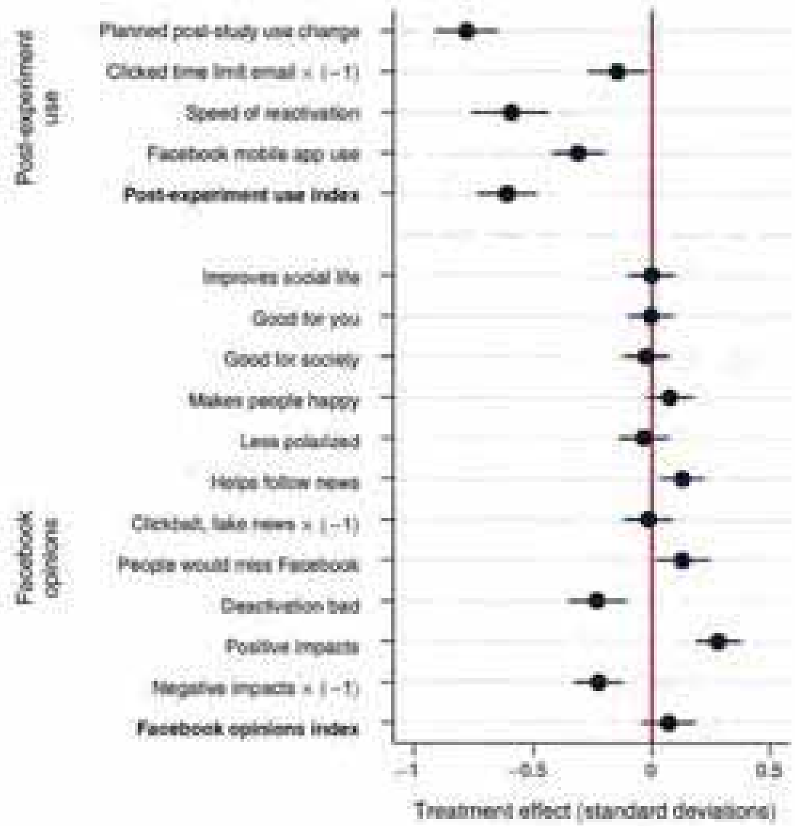


Figure 6. Effects on Post-Experiment Facebook Use and Opinions

Notes: This figure presents local average treatment effects of Facebook deactivation estimated using equation (1). All variables are normalized so that the Control group endline distribution has a standard deviation of 1. Error bars reflect 95 percent confidence intervals. See Section IC for variable definitions.

Others described the difficulty of not being able to post for special events such as family birthdays and not being able to participate in online groups.

Overall, our data suggest that Facebook does indeed have adverse effects on SWB. However, the magnitude of these effects is moderate and may be smaller than correlation studies would suggest, and our qualitative interviews suggest that the average effect likely masks substantial heterogeneity.

D. *Post-Experiment Facebook Use and Opinions*

Figure 6 presents effects of deactivation on post-experiment demand for Facebook as well as participants' subjective opinions about Facebook. These results are closely related to the findings on subjective well-being, as we might expect participants who found deactivation increased their happiness would choose to use Facebook less in the future. They also speak more directly to the popular debate over whether social media are addictive and harmful. If deactivation

reduces post-experiment Facebook use, this is consistent with standard habit formation models such as Becker and Murphy (1988), or with learning models in which experiencing deactivation caused people to learn that they would be better off if they used Facebook less.[28]

Deactivation clearly reduced post-experiment demand for Facebook. These effects are very stark, with by far the largest magnitude of any of our main findings. The effect on reported intentions to use Facebook as of the endline survey is a reduction of 0.78 standard deviations: while the average Control group participant planned to reduce future Facebook use by 22 percent, deactivation caused the Treatment group to plan to reduce Facebook use by an additional 21 percent relative to Control. In our post-endline survey a month after the experiment ended, we measured whether people actually followed through on these intentions, by asking people how much time they had spent on the Facebook mobile app on the average day in the past week. Deactivation reduces this post-endline Facebook mobile app use by 12 minutes per day, or 0.31 standard deviations. This is a 23 percent reduction relative to the Control group mean of 53 minutes per day, lining up almost exactly with the planned reductions reported at endline. However, online Appendix Table A13 shows that the reduction is less than half as large (8 percent of the Control group mean) and not statistically significant (with a t-statistic of -1.16) if we limit the sample to iPhone users who reported their usage as recorded by their Settings app, thereby excluding participants who were reporting personal estimates of their usage.

As a different (and non-self-reported) measure of post-experiment use, we can look at the speed with which people reactivated their Facebook accounts following the 24-hour post-endline period in which both Control and Treatment were deactivated. Figure 7 presents the share of our deactivation checks in which the Treatment and Control groups were deactivated, by day of the experiment.[29] By day 35, one week after the end of the experiment, 11 percent of the Treatment group was still deactivated, compared to 3 percent of the Control group. By day 91, nine weeks after the end of the experiment, 5 percent of the Treatment group was still deactivated, against 2.5 percent of Control. As Figure 6 shows, the local average treatment effect on the speed of reactivation is a highly significant 0.59 standard deviations. Overall, deactivation clearly decreased post-experiment use, reducing the index by 0.61 standard deviations. As introduced above, this is consistent with models of habit formation or learning.

The bottom group of outcomes in Figure 6 supplement the post-experiment use outcomes by measuring participants' qualitative opinions about Facebook. These are re-signed so that more positive means more positive opinions, so agreement with the statement that "Facebook exposes people to clickbait or false news stories" and the length of text about Facebook's negative impacts are both multiplied by (-1).

[28] Online Appendix Figure A33 presents histograms of participants' opinions about Facebook at baseline. People are evenly divided on whether Facebook is good or bad for themselves and for society and whether Facebook makes people more or less happy. Consistent with our results, people tend to think that Facebook helps people to follow the news better and makes people more politically polarized.

[29] There is a slight dip in deactivation rates for the Treatment group seven days after the deactivation period began. This was caused by the fact that some participants failed to turn off a default setting in which Facebook reactivates users' profiles after seven days of deactivation. For technical reasons, our deactivation checking algorithm checked the entire Control group once every few days between midline and endline in order to check the Treatment group four times per day. After endline, we returned to checking all participants approximately once per day.

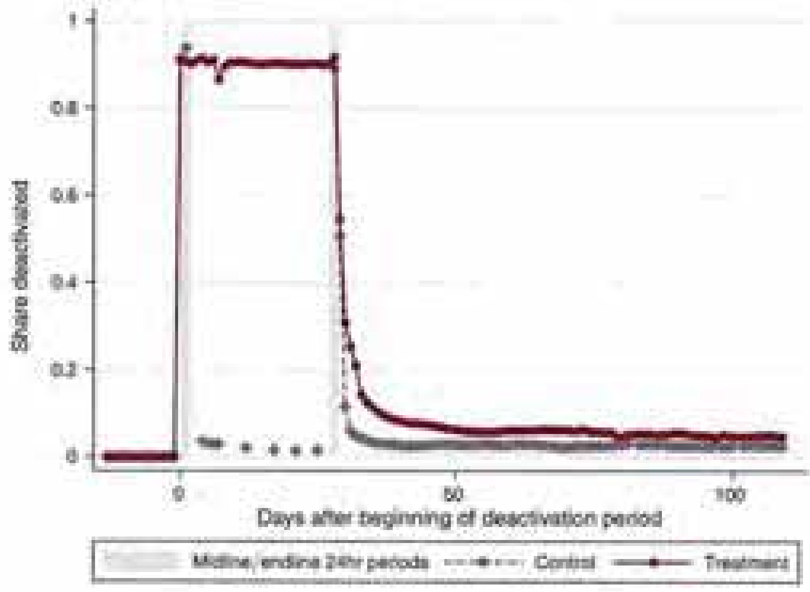


Figure 7. Probability of Being Deactivated

Notes: This figure shows the share of the Treatment and Control groups that had their Facebook accounts deactivated, by day of the experiment, for the impact evaluation sample: participants who were willing to accept less than \$102 to deactivate Facebook for the four weeks after midline and were offered $p = \$102$ or $p = \$0$ to do so. The vertical gray areas reflect the 24-hour periods after midline and endline during which both Treatment and Control were instructed to deactivate.

The results are mixed. Deactivation increases the extent to which participants think Facebook helps them follow the news better, and it also makes participants agree more that people would miss Facebook if they stopped using it. On the other hand, participants who deactivated for four weeks instead of 24 hours were more likely to say that their deactivation was good for them.[30] Deactivation increases both the *positive impacts* and *negative impacts* variables, i.e., it makes people write more about both positive and negative aspects of Facebook. Overall, deactivation had no statistically significant effect on the Facebook opinions index.

Figure 8 presents the distributions of Treatment and Control responses to two key questions reflecting opinions about Facebook. Both Treatment and Control tended to agree that "if people spent less time on Facebook, they would soon realize that they don't miss it," but deactivation weakened that view. On this figure, the Treatment group's average response on the scale from -5 to $+5$ was -1.8, while the Control group's average response is -2.0. The right panel shows that both Treatment and Control tended to think that deactivation was good for them, but the Treatment group is more likely to think that their (longer) deactivation was good for them. On this figure, the Treatment group's average response on the scale from

[30]One should be cautious in interpreting this effect, as it could result both from a change of opinion about Facebook and from the difference in length of the deactivation they were evaluating. As we shall see below, the Control group also tends to believe that deactivation was good for them, but the modal answer was 0 (i.e., neither good nor bad), suggesting that many people were indifferent to such a short deactivation.

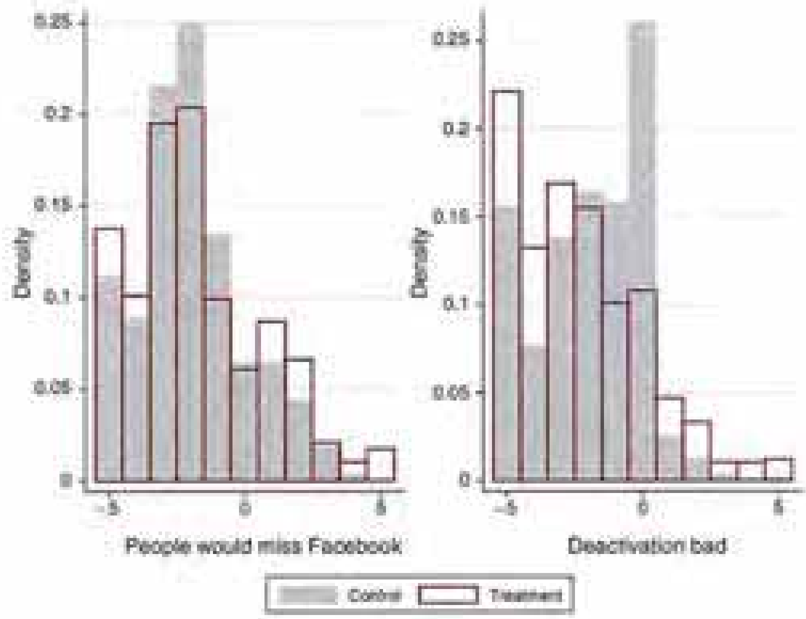


FIGURE 8. KEY OPINIONS ABOUT FACEBOOK IN TREATMENT AND CONTROL

Notes: This figure presents the distribution of responses in Treatment and Control for two key measures of opinions about Facebook. See Section IC for variable definitions.

−5 to +5 is −2.3, while the Control group's average response is −1.9. Remarkably, about 80 percent of the Treatment group thought that deactivation was at least somewhat good for them, and the modal response was the strongest possible agreement that deactivation was good (the left-most bar on the histogram). In both panels, the Treatment group has a wider dispersion of responses, with more people strongly agreeing *and* more people strongly disagreeing. This highlights the importance of testing for treatment effect heterogeneity, which we will do in the next section.

To give a richer sense of how deactivation affected Facebook use, the post-endline survey included a free-response question in which we asked people to write how they had changed their Facebook use since participating in the study. We then use standard text analysis tools to determine how the Treatment and Control groups responded differently. Specifically, we processed the text by stemming words to their linguistic roots (for example, "changes," "changing," and "changed" all become "chang"), removing common "stop words" (such as "the" and "that"), and making lists of all one-, two-, three-, and four-word phrases that appeared five or more times in the sample. We then constructed Pearson's χ^2 statistic, which measures the extent of differential usage rates between Treatment and Control; the phrases with the highest χ^2 are especially unbalanced between the two groups. This parallels Gentzkow and Shapiro's (2011) approach to determining which phrases are used more by Republicans versus Democrats, except we determine which phrases are used more by Treatment versus Control.

The two panels of Table 4 present the 20 highest-χ^2 phrases that were more common in Treatment and in Control. The Treatment group was relatively likely

TABLE 4—MOST COMMON DESCRIPTIONS OF FACEBOOK USE CHANGES

Phrases used more often by Treatment			Phrases used more often by Control		
Phrase	% Treatment	% Control	Phrase	% Treatment	% Control
Not use facebook anymor	0.90	0	Ha not chang	6.63	16.76
Not spend much time	1.08	0.36	Not chang sinc particip	0	0.99
Spend less time facebook	0.90	0.27	Ha not chang sinc	0.18	1.53
Have not use facebook	0.72	0.18	Chang sinc particip studi	0	0.81
Not use facebook much	0.72	0.18	Way use facebook ha	0.18	1.35
Spend lot less time	0.72	0.27	Usag ha not chang	0	0.72
Use much less	2.87	0.63	Chang way use facebook	0.18	1.26
Definit use facebook	0.54	0.18	Not chang	7.17	18.65
Use facebook lot less	0.54	0.18	Awar much time spend	0	0.63
Use facebook much less	0.54	0.18	Ha not	8.24	19.64
Not use facebook	3.05	1.17	Not much ha chang	0	0.54
Use litl bit less	0.54	0.27	Way use facebook	0.54	2.70
Have not use	1.25	0.18	Not think chang much	0	0.45
Ha not chang use	0.72	0.45	Not chang much use	0	0.45
Use facebook anymor	0.90	0.09	Use facebook slightli less	0	0.45
Think use less	1.61	0.45	More awar much time	0.18	0.99
No ha not chang	0.54	0.36	Chang sinc particip	0	1.08
Use news app	0.72	0.09	Much time spend	0.18	1.53
Still have not	0.90	0.18	Facebook ha not chang	0.72	2.07
Much less	4.84	1.17	Use slightli less	0	0.90

Notes: The post-endline survey included the following question with an open response text box: "How has the way you use Facebook changed, if at all, since participating in this study?" For all responses, we stemmed words, filtered out stop words, then constructed all phrases of length $l = \{1,2,3,4\}$ words. For each phrase p of length l, we calculated the number of occurrences of that phrase in Treatment and Control group responses (f_{plT} and f_{plC}) and the number of occurrences of length-l phrases that are *not* phrase p in Treatment and Control responses ($f_{\sim plT}$ and $f_{\sim plC}$). We then constructed Pearson's χ^2-statistic:

$$\chi^2 = \frac{(f_{plT}f_{\sim plC} - f_{plC}f_{\sim plT})^2}{(f_{plT}+f_{plC})(f_{plT}+f_{\sim plT})(f_{plC}+f_{\sim plC})(f_{\sim plT}+f_{\sim plC})}.$$

This table presents the 20 phrases with the highest χ^2 that were most commonly written by the Treatment and Control groups. The % Treatment and % Control columns present the share of people in the respective group whose responses included each phrase.

to write that they were using Facebook less or not at all ("use much less," "not use facebook anymor," "stop use facebook") or more judiciously: the phrase "use news app" is mostly from people saying that they have switched to getting news from their phone's news app instead of Facebook. By contrast, while a few of the Control group's most common phrases indicate lower use (variants of "more aware much time spend" and "use facebook slightli less"), the great majority of their relatively common phrases indicate that their Facebook use has not changed.

To more deeply understand the ways in which deactivation changed people's relationship to Facebook, we partnered with a team of qualitative researchers who analyzed our survey data and additional participant interviews (Baym, Wagman, and Persaud forthcoming). They find that many participants emphasized that their time off of Facebook led them to use the platform more "consciously," aligning their behavior with their desired use. For example, some participants discussed avoiding their news feed and only looking at their Facebook groups, while others removed the Facebook app from their phones and only accessed the site using their computers.

E. *Heterogeneous Treatment Effects*

Individual Moderators.—In our pre-analysis plan, we specified that we would present separate estimates for subgroups defined by four primary moderators. Figure 9 presents those estimates. The top panel presents estimates for *heavy users* versus *light users*: that is, people whose baseline reported Facebook use was above versus below median. There is no consistent evidence that the effects are different for people who report being heavier users, perhaps because Facebook use is measured with noise.

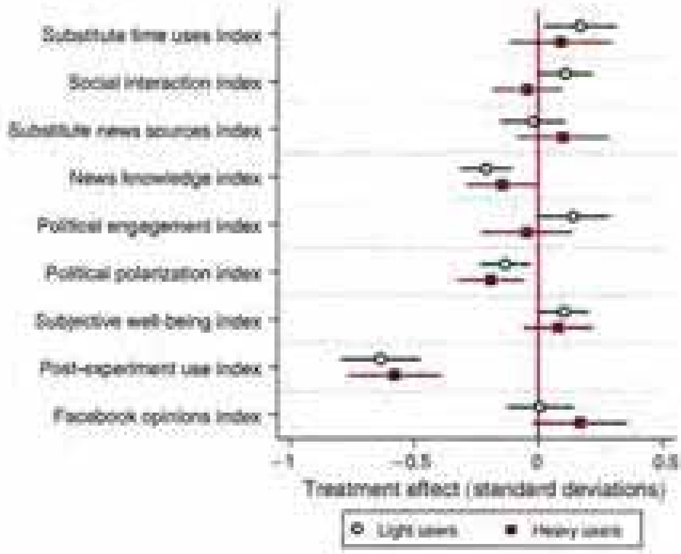
The second panel presents estimates for *heavy news users* versus *light news users*: that is, those who get news from Facebook fairly often or very often versus never, hardly ever, or sometimes. As one might expect, the estimated effects for news knowledge are larger for people who get more news from Facebook, but this difference is not statistically significant. The pre-analysis plan specified that we would limit these tests to only the news and political outcomes in Section IVB.

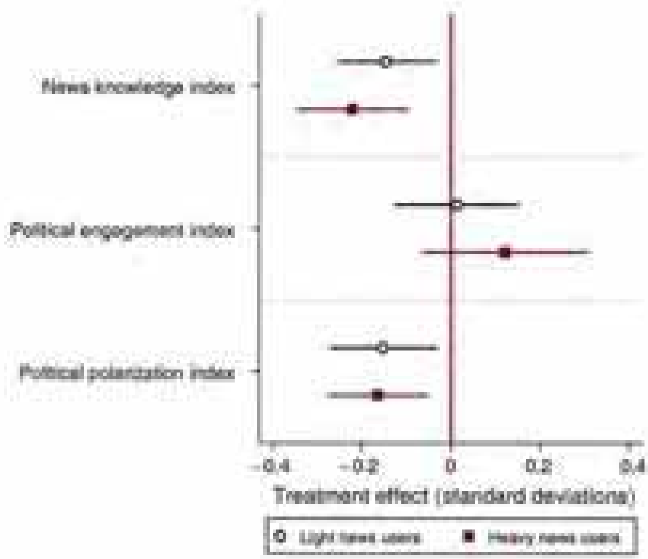
The third panel presents separate estimates for *active users* versus *passive users*. We measure this using two questions: share of active versus passive browsing using a question based on the Passive and Active Facebook Use Measure (Gerson, Plagnol, and Corr 2017), and "what share of your time on Facebook do you spend interacting one-on-one with people you care about." Active versus passive users are defined as having above- versus below-median sum of their two responses to these questions. This moderator is of interest because of a set of papers cited in the introduction suggesting that passive Facebook use can be harmful to subjective well-being, while active use might be neutral or beneficial. Perhaps surprisingly, we see no differences in the effects of deactivation on the subjective well-being index. The pre-analysis plan specified that we would limit these tests to the four families reported in the figure.

Finally, the fourth panel presents separate estimates of effects on subjective well-being text message surveys for text messages sent during the time of day when the respondent reported using Facebook the most. We see no clear differences in the effects on subjective well-being.

The pre-analysis plan also specified two secondary moderators: age (for all outcomes) and political party (limited to the news and political outcomes). We considered these secondary because we did not have a strong prior that we would be able to detect heterogeneous effects. Online Appendix Figure A9 presents estimates of effects on these outcomes. There are no systematic patterns.

Online Appendix Figure A9 also includes heterogeneity by above- versus below-median valuation of Facebook. While we added this moderator only after the pre-analysis plan was submitted, it is important because our impact evaluation sample only includes participants with WTA less than \$102. Under the assumption that marginal treatment effects are monotonic in WTA, treatment effect heterogeneity within our impact evaluation sample would be informative about treatment effects for the full population. The effects for above- versus below-median WTA differ statistically for only one index: the effects on political polarization are driven by above-median WTA participants. The above-median WTA point estimate is larger and statistically indistinguishable for two indices, smaller and statistically





(Continued)

FIGURE 9. HETEROGENEOUS TREATMENT EFFECTS

indistinguishable for four indices, and opposite-signed for the final index. This provides some support for the view that effect sizes would not be systematically different in the full Facebook user population including users with higher valuations.

Online Appendix Figure A9 presents one additional test of external validity that was suggested by a referee after the pre-analysis plan was submitted. We construct sample weights that match the impact evaluation sample to the observable characteristics of Facebook users in Table 2. Online Appendix Figure A9 shows that participants with below- versus above-median sample weights, that is, the types of

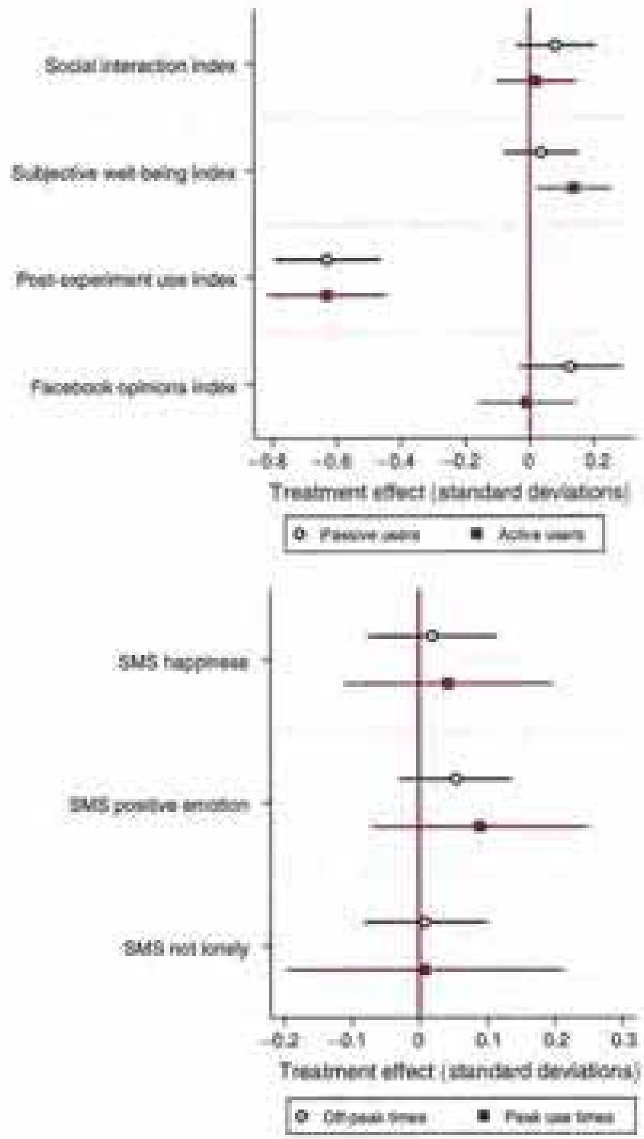


Figure 9. Heterogeneous Treatment Effects (Continued)

Notes: This figure presents local average treatment effects of Facebook deactivation estimated using equation (1), for subgroups defined by the primary moderators in our pre-analysis plan. All variables are normalized so that the Control group endline distribution has a standard deviation of 1. Error bars reflect 95 percent confidence intervals. See Section IC for variable definitions.

people who were especially likely versus unlikely to participate in the study, do not have systematically different treatment effects. This provides some further support for the view that effect sizes would be similar in the full Facebook user population.

Online Appendix F presents heterogeneous treatment effects on each individual outcome.

All Possible Moderators.—Many factors other than the specific variables we specified above might moderate treatment effects of Facebook deactivation. To search for additional possible moderators, we test whether any of the demographics or outcome variable indices collected at baseline might moderate treatment effects on the key outcomes of interest. We consider six outcomes: the latter five indices (news knowledge, political polarization, subjective well-being, post-experiment use, and Facebook opinions) plus the variable *Deactivation bad*, which we add because of the heterogeneity displayed in Figure 8. We consider 13 potential moderators: all 6 demographic variables listed in Table 2 (income, years of education, gender, race, age, and political party affiliation, which is on a seven-point scale from strongly Democratic to strongly Republican) and the baseline values of all 7 relevant indices.[31] We normalize each potential moderator to have a standard deviation of 1, and we denote normalized moderator k by X_i^k.

For all outcomes other than *Deactivation bad*, we estimate the following modified version of equation (1):

$$Y_i = \tau D_i + \alpha^k D_i X_i^k + \zeta X_i^k + \rho \mathbf{Y}_i^b + \nu_s + \varepsilon_i, \tag{4}$$

instrumenting for D_i and $D_i X_i^k$ with T_i and $T_i X_i^k$. For *Deactivation bad*, we simply estimate $Y_i = \alpha^k X_i^k + \varepsilon_i$ in the Treatment group only; this identifies what types of people in the Treatment group thought that deactivation was particularly good or bad. In total, we carry out 78 tests in 78 separate regressions: 13 potential moderators for each of the 6 outcomes.

There are many ways to estimate heterogeneous treatment effects, including causal forests (Athey, Tibshirani, and Wagner 2019) and lasso procedures. We chose this approach because it delivers easily interpretable estimates.

Figure 10 presents the interaction coefficients $\hat{\alpha}^k$ and 95 percent confidence intervals for each of the six outcomes. To keep the figures concise, we plot only the five moderators with the largest absolute values of $\hat{\alpha}^k$, so there are another eight smaller unreported $\hat{\alpha}^k$ coefficients for each outcome.

We highlight three key results. First, deactivation may reduce polarization more (i.e., Facebook use may increase polarization more) for older people, white people, and men. Second, Facebook deactivation has less positive effect on subjective well-being for people who have more offline social interactions and are already more happy at baseline. This suggests that Facebook use may have the unfortunate effect of reducing SWB more for people with greater social and psychological need. In our sample, these "higher-need" people also use Facebook more heavily. Third, people may have some intuition about whether they will like deactivation: people with more positive baseline opinions about Facebook are less likely to decrease their post-experiment use and less likely to think that deactivation was good for them.

[31] There are originally nine indices. We exclude the baseline substitute time uses index because it is not easily interpretable, and we exclude the baseline post-experiment use index because this only includes *Facebook mobile app use*.

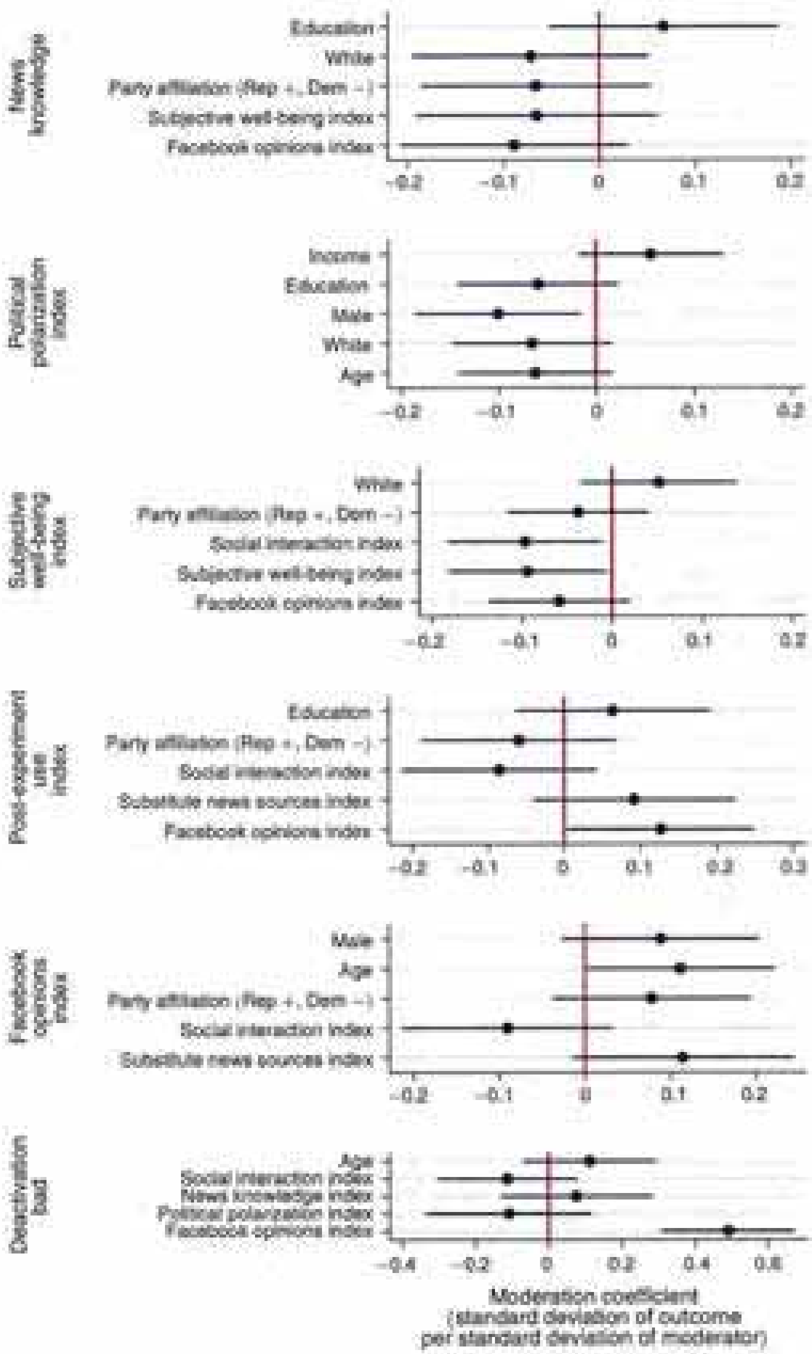


FIGURE 10. HETEROGENEOUS TREATMENT EFFECTS FOR ALL MODERATORS

Notes: This figure presents the moderators of local average treatment effects of Facebook deactivation estimated using equation (4). For each of the six outcomes, we present the five moderators with the largest moderation coefficients $\hat{\alpha}^k$. All outcome variables are normalized so that the Control group endline distribution has a standard deviation of 1, and all moderators are also normalized to have a standard deviation of 1. Error bars reflect 95 percent confidence intervals. See Section IC for variable definitions.

TABLE 5—PERCEIVED RESEARCHER AGENDA IN TREATMENT AND CONTROL

Variable	Treatment mean/SD (1)	Control mean/SD (2)	t-test p-value (1) − (2)
I don't think they had a particular agenda	0.43 (0.49)	0.44 (0.50)	0.59
Yes, wanted to show that Facebook is good for people	0.03 (0.18)	0.04 (0.19)	0.79
Yes, wanted to show that Facebook is bad for people	0.35 (0.48)	0.35 (0.48)	0.79
I am not sure	0.19 (0.39)	0.18 (0.38)	0.62
Observations	573	1,064	

Notes: The endline survey asked, "Do you think the researchers in this study had an agenda?" Columns 1 and 2 present the share of the Treatment and Control groups who gave each possible response. Column 3 presents p-values of tests of differences in means between the two groups.

F. *Experimenter Demand Effects*

Most of our outcomes are self-reported, and it would have been difficult to further conceal the intent of the randomized experiment. This raises the possibility of experimenter demand effects, i.e., that survey responses depend on what participants think the researchers want them to say. To test for demand effects, the endline survey asked, "Do you think the researchers in this study had an agenda?" Table 5 presents the possible responses and shares by treatment group.

For demand effects to arise, participants must believe that the researchers want a particular pattern of responses. Table 5 shows that 62 percent of both Treatment and Control groups thought we had no particular agenda or were not sure. This suggests that demand effects would not arise for a solid majority of our sample. However, demand effects could arise for the remaining 38 percent.

For experimenter demand effects to bias our treatment effects, either (i) the Treatment and Control groups must have different beliefs about what the researchers want, or (ii) participants must sense what treatment group they are in and change their answers to generate the treatment effect that they think the researchers want (or don't want). Table 5 shows that possibility (i) is not true: perceived researcher agenda is closely balanced between Treatment and Control. To test for possibility (ii), we can estimate treatment effects separately for the subsample that thought that we "wanted to show that Facebook is bad for people" versus all other participants. If (ii) is true, then our results should be different in these two subsamples. Online Appendix Figure A36 shows that this is not the case: the effects on outcome indices that look "good" or "bad" for Facebook (e.g., news knowledge, political polarization, subjective well-being, and post-experiment use) are not statistically different, and there is no pattern of point estimates to suggest that the results are generally more "good" or "bad" in one of the two subsamples.

Of course, these tests are only suggestive. But combined with the fact that the non-self-reported outcomes paint a similar picture to the self-reports, these tests suggest that demand effects are unlikely to be a major source of bias in our results.

V. Measuring the Consumer Surplus from Facebook

Quantifying the economic gains from free online services such as search and media is particularly important given that these services represent an increasingly large share of the global economy. This measurement has been particularly challenging because the lack of price variation (or any price at all) makes it impossible to use standard demand estimation to measure consumer surplus.[32] In this section, we present two back-of-the-envelope consumer surplus calculations. First, we employ the standard assumption that willingness-to-accept identifies consumer surplus. Second, we adjust consumer surplus to account for the possibility that deactivation might help people learn their true valuation of Facebook. This adjustment highlights the challenges in using willingness-to-accept as a measure of consumer welfare.

A. *Standard Consumer Surplus Estimate*

In a standard model, willingness-to-accept to abstain from Facebook equals consumer surplus. Figure 11 presents the histogram of WTA to deactivate Facebook for the four weeks after midline instead of only the 24 hours after midline. The median is \$100, and almost 20 percent had valuations greater than \$500. After winsorizing valuations at \$1,000, the mean is \$203. After reweighting the sample to match the observable characteristics of Facebook users in Table 2, the median is still \$100, and the winsorized mean is \$180. Multiplying the mean by the estimated 172 million US Facebook users would imply that 27 days of Facebook generates \$31 billion of consumer surplus.

Our sample's WTA for Facebook abstention is larger than in most other studies, but not all. In an online panel weighted for national representativeness, Brynjolfsson, Eggers, and Gannamaneni (2018) estimates that the mean WTA to not use Facebook for one month is \$48, and that the median WTA to hypothetically stop using social media for one year was \$205 in 2016 and \$322 in 2017. In their sample of European college students, Brynjolfsson, Eggers, and Gannamaneni (2018) finds a median WTA of \$175 for one month.[33] In samples of college students, residents of a college town, and Amazon MTurk workers, Corrigan et al. (2018) estimates that the mean annualized WTA to deactivate Facebook ranges from \$1,139 to \$1,921, depending on the sample and the length of deactivation. In a sample of college students, Mosquera et al. (2018) estimates that the median (mean) WTA to not use Facebook for one week is \$15 (\$25). In an unincentivized (stated preference) survey of MTurk workers, Sunstein (forthcoming) found a \$1 per month median willingness-to-pay for Facebook and a \$59 per month median willingness-to-accept to not use Facebook.

There are many caveats to using this type of stylized calculation to approximate the consumer surplus from Facebook. First, we (and Corrigan et al.) required participants to deactivate their Facebook accounts instead of simply abstaining from logging in. For people who planned to avoid using other apps with Facebook logins

[32] As mentioned in the introduction, see Brynjolfsson and Saunders (2009); Byrne, Fernald, and Reinsdorf (2016); Nakamura, Samuels, and Soloveichik (2016); Brynjolfsson, Rock, and Syverson (2019); and Syverson (2017).

[33] Online Appendix Figure A37 compares our demand curve to the Brynjolfsson, Eggers, and Gannamaneni (2018) demand curves.

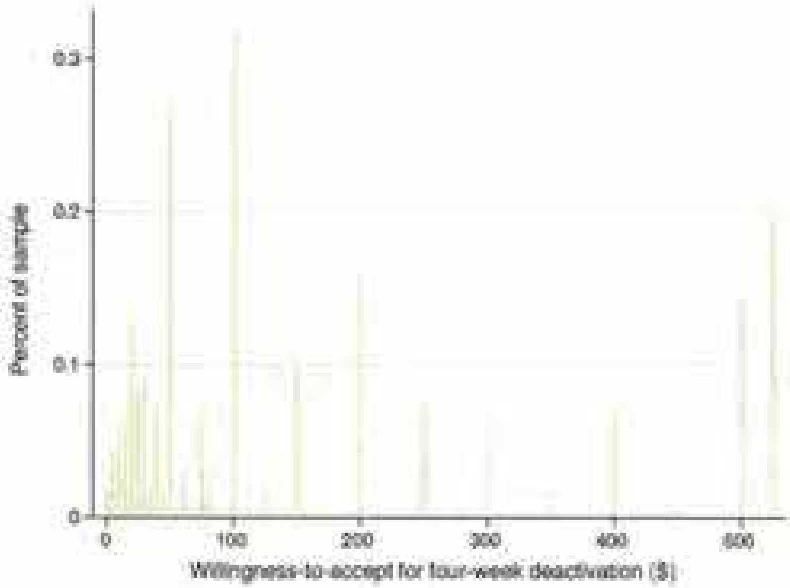


Figure 11. Distribution of Willingness-to-Accept to Deactivate Facebook after Midline

Notes: This figure presents the distribution of willingness-to-accept to deactivate Facebook between midline and endline. All responses above $525 are plotted at $525.

in order to avoid reactivating their Facebook accounts, WTA overstates the value of Facebook access. Second, participants must believe the experimenter will in fact enforce deactivation; WTA could naturally be lower for a partially enforced or unenforced deactivation compared to an enforced deactivation. In some other studies, the method of enforcement was either not made clear ex ante, or enforcement was not fully carried out ex post.[34] Third, any survey sample is unlikely to be representative of the Facebook user population on both observable and unobservable characteristics. For example, we screened out people who reported using Facebook 15 minutes or less per day, and while we reweight the average WTAs to match the average observables of Facebook users (including average daily usage), this reweighting may implicitly overstate the WTA of people who don't use Facebook very much. Fourth, we (and all other existing studies) estimate people's Facebook valuations holding their networks fixed. Due to network externalities, valuations could be quite different if participants' friends and family also deactivated. Fifth, one should be careful in annualizing these estimates or comparing WTAs for different durations of abstention, as our study and several others find that the average per-day valuation varies with the duration. Sixth, as we will see, in practice people's WTA may not be closely held and could be easily anchored or manipulated, even in incentive

[34] Mosquera et al. told participants that they would "require" that they "not use their Facebook accounts" but did not give additional details. Brynjolfsson et al.'s WTA elicitation stated that the experimenters "will randomly pick 1 out of every 200 respondents and her/his selection will be fulfilled," and that they could enforce deactivation by observing subjects' time of last login, "given your permission." In practice, the deactivation was mostly not enforced: of the ten subjects randomly selected for enforcement, one gave permission.

compatible elicitations such as ours. Finally, this calculation fails to speak to the possibility that people misperceive Facebook's value. We turn to that issue now.

B. *How Deactivation Affects Valuations*

It is often argued that social media users do not correctly perceive the ways in which social media could be addictive or make them unhappy. If this is the case, people's willingness-to-accept to abstain from Facebook would overstate "true" consumer surplus. For example, Alter (2018), Newport (2019), many popular media articles,[35] and organizations such as the Center for Humane Technology and Time to Log Off argue that Facebook and other digital technologies can be harmful and addictive. The Time to Log Off website argues that "everyone is spending too much time on their screens" and runs "digital detox campaigns." Sagioglu and Greitemeyer (2014) documents an "affective forecasting error": people predicted that spending 20 minutes on Facebook would make them feel better, but a treatment group randomly assigned to 20 minutes of Facebook browsing actually reported feeling worse.

Some of our results are also consistent with this argument. In the baseline survey, two-thirds of people agreed at least somewhat that "if people spent less time on Facebook, they would soon realize that they don't miss it." As reported earlier, about 80 percent of the Treatment group thought that deactivation was good for them, and both qualitative and quantitative data suggest that deactivation caused people to rethink and reoptimize their use.

The core of this argument is that people's social media use does not maximize their utility, and a "digital detox" might help them align social media demand with their own best interests. This idea is related to several existing economic models. In a model of projection bias (Loewenstein, O'Donoghue, and Rabin 2003), people might not correctly perceive that social media are habit forming or that their preferences might otherwise change after a "digital detox." In an experience good model, a "digital detox" might help consumers to learn their valuation of social media relative to other uses of time. Of course, both of these mechanisms could also affect demand after a period of deactivation, so it is not clear whether the WTA before deactivation or after deactivation is more normatively relevant.

To provide evidence on these issues, we elicited WTA at three points, as described earlier. First, on the midline survey, we elicited WTA to deactivate Facebook in "weeks 1–4" (the four weeks after midline). We call this WTA w_1. Second, just after telling people their BDM offer price on the midline survey, and thus whether they were expected to deactivate in weeks 1–4, we elicited WTA to deactivate in "weeks 5–8" (the four weeks after endline). We call this $w_{2,1}$. Third, on the endline survey, we elicited WTA to deactivate in weeks 5–8, after the Treatment group had experienced deactivation in weeks 1–4, but the Control group had not. We call this $w_{2,2}$.

[35] For example: Chris Ciaccia, "Facebook, Cocaine, Opioids: How Addictive Is the Social Network?" *Fox News*, December 29, 2017, https://www.foxnews.com/tech/facebook-cocaine-opioids-how-addictive-is-the-social-network; Will Oremus, "Addiction for Fun and Profit," *Slate*, November 10, 2017, https://slate.com/technology/2017/11/facebook-was-designed-to-be-addictive-does-that-make-it-evil.html.

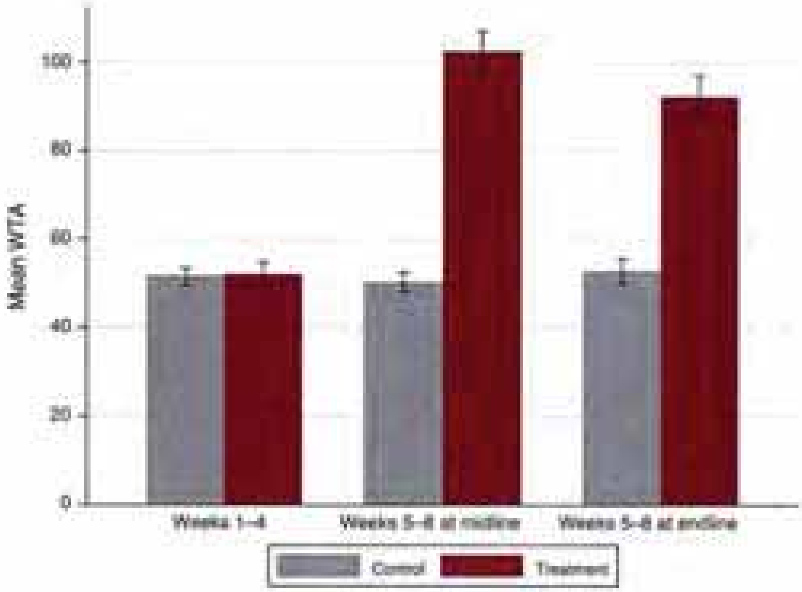


FIGURE 12. AVERAGE VALUATION OF FACEBOOK IN TREATMENT AND CONTROL

Notes: This figure presents the mean willingness-to-accept (WTA) to deactivate Facebook in Treatment and Control, for the impact evaluation sample: participants who were willing to accept less than \$102 to deactivate Facebook for the four weeks after midline and were offered $p = \$102$ or $p = \$0$ to do so. The first pair of bars is the mean WTA for deactivation in weeks 1–4, the four weeks after the midline survey. The second pair of bars is mean WTA for deactivation in weeks 5–8, the four weeks after the endline survey, as elicited in the midline survey. The third pair of bars is mean WTA for deactivation in weeks 5–8, as elicited in the endline survey.

The Control group's change in WTA for weeks 5–8, $\Delta w_2 = w_{2,2} - w_{2,1}$, captures any unpredicted time effect. The Treatment group's WTA change Δw_2 reflects both the time effect and the unexpected change in valuation caused by deactivation. If the time effect is the same in both groups, then the difference-in-differences measures the effect of deactivation on valuations due to projection bias, learning, and similar mechanisms.

Figure 12 presents the average of WTA in Treatment and Control of w_1, $w_{2,1}$, and $w_{2,2}$. Recall that the impact evaluation sample includes only people with $w_1 < \$102$, so these averages are less than the unconditional means discussed above and presented in Figure 11. Because of outliers in the WTAs for weeks 5–8, we must winsorize WTA. We winsorize at \$170 for this figure and our primary regression estimates, as this is the upper bound of the distribution of BDM offers that we actually made for deactivation.

The Treatment group's valuation for weeks 5–8 jumps substantially relative to its valuation for weeks 1–4, while the Control group's valuation for weeks 5–8 does not. We used open-answer questions in the post-endline survey and qualitative interviews to understand this change. Some of the large gap may be due to costs of deactivation being convex in the length of deactivation: some people in the Treatment group wrote that they were much less comfortable deactivating for eight weeks instead of four, as they would have to make much more extensive arrangements to communicate with friends, coworkers, and schoolmates during a longer deactivation. However, participants' open-answer responses suggest that the

Treatment group's WTA increase is also affected by anchoring on the \$102 BDM offer that was revealed after the elicitation of w_1 but before the elicitation of $w_{2,1}$. Such anchoring is consistent with prior results showing that valuations elicited using the BDM method can be affected by suggested prices or other anchors (Bohm, Lindén, and Sonnegård 1997; Mazar, Kőszegi, and Ariely 2014). Thus, we do not believe this increase is relevant for a consumer welfare calculation, and we do not draw any substantive conclusion from it.

Figure 12 also illustrates Δw_2, the change in valuation of weeks 5–8 between midline and endline. The Control group's valuation increases, reflecting an unpredicted time effect. In open-answer questions, some people wrote that they were less willing to deactivate during the Thanksgiving holiday, and they may not have foreseen this as of the midline survey on October 11. By contrast, the Treatment group's valuation for weeks 5–8 decreases. Thus, the difference-in-differences Δw_2 is negative.

We can estimate the difference-in-differences using the following regression:

$$\Delta w_{2,i} = \gamma D_i + \rho w_{1,i} + \nu_s + \varepsilon_i, \tag{5}$$

instrumenting for D_i with T_i. Table 6 presents results, winsorizing all WTAs at \$170 in column 1 and at \$1,000 in column 2. Relative to the Control group, the Treatment group reduced its post-endline valuation by \$14 to \$18, or about 14 percent of the Treatment group's average $w_{2,1}$. This suggests that deactivation eliminated projection bias or facilitated learning that reduced demand for Facebook by 14 percent. In turn, this suggests that the traditional estimates might somewhat overstate consumer surplus.

This result is consistent with our finding in Section IVD that deactivation reduced post-experiment Facebook use. However, because the WTA update Δw_2 is unexpected, it suggests that the results from Section IVD may not be entirely explained by a "rational" habit formation model such as Becker and Murphy (1988), in which people foresee how consumption affects future marginal utility. Instead, these results suggest that at least some of the reduced Facebook demand caused by deactivation is driven by unexpected factors such as projection bias and learning.

One caveat is that the anchoring effect described above could affect our estimate of γ. If anchoring has the same effects on $w_{2,1}$ and $w_{2,2}$ in the Treatment group, then Δw_2 is unaffected, and our estimate of γ is unbiased. If the anchoring effects decay between midline and endline, this would bias $\hat{\gamma}$ away from zero, meaning that the true γ would be less than our estimate.[36] This would further strengthen our result that the valuation update caused by deactivation equals only a small share of valuations.

One interpretation of these results is that they reinforce the standard model calculation that Facebook generates many billions of dollars in consumer surplus. Another interpretation is that they further highlight why standard consumer surplus calculations based on elicited valuations can be problematic.

[36] An alternative experimental design choice we considered was to elicit $w_{2,1}$ *before* revealing the weeks 1–4 offer price, separately for the case in which the participant would be paid to deactivate for weeks 1–4 and the case in which the participant would not be paid to deactivate. In this case, however, any anchoring effect would have appeared on $w_{2,2}$ but not $w_{2,1}$, generating an unambiguous spurious treatment effect on Δw_2.

TABLE 6—CHANGE IN FACEBOOK VALUATION AFTER DEACTIVATION

	(1)	(2)
Share of time deactivated	−14.36	−18.22
	(2.60)	(7.73)
Observations	1,634	1,634
Winsorized maximum WTA	170	1,000
Treatment mean weeks 5–8 WTA at midline	103	135

Notes: This table presents estimates of equation (3). The dependent variable is the change in WTA for post-endline deactivation measured at endline versus midline. Standard errors are in parentheses.

VI. Conclusion

Our results leave little doubt that Facebook provides large benefits for its users. Even after a four-week "detox," our participants spent substantial time on Facebook every day and needed to be paid large amounts of money to give up Facebook. Our results on news consumption and knowledge suggest that Facebook is an important source of news and information. Our participants' answers in free response questions and follow-up interviews make clear the diverse ways in which Facebook can improve people's lives, whether as a source of entertainment, a means to organize a charity or an activist group, or a vital social lifeline for those who are otherwise isolated. Any discussion of social media's downsides should not obscure the basic fact that it fulfills deep and widespread needs.

Notwithstanding, our results also make clear that the downsides are real. We find that four weeks without Facebook improves subjective well-being and substantially reduces post-experiment demand, suggesting that forces such as addiction and projection bias may cause people to use Facebook more than they otherwise would. We find that while deactivation makes people less informed, it also makes them less polarized by at least some measures, consistent with the concern that social media have played some role in the recent rise of polarization in the United States. The estimated magnitudes imply that these negative effects are large enough to be real concerns, but also smaller in many cases than what one might have expected given prior research and popular discussion.

The trajectory of views on social media, with early optimism about great benefits giving way to alarm about possible harms, is a familiar one. Innovations from novels to TV to nuclear energy have had similar trajectories. Along with the important existing work by other researchers, we hope that our analysis can help move the discussion from simplistic caricatures to hard evidence, and provide a sober assessment of the way a new technology affects both individual people and larger social institutions.

REFERENCES

►**Acland, Dan, and Matthew Levy.** 2015. "Naiveté, Projection Bias, and Habit Formation in Gym Attendance." *Management Science* 61 (1): 146–60.

►**Allcott, Hunt, and Matthew Gentzkow.** 2017. "Social Media and Fake News in the 2016 Election." *Journal of Economic Perspectives* 31 (2): 211–36.

Alter, Adam. 2018. *Irresistible: The Rise of Addictive Technology and the Business of Keeping Us Hooked.* New York: Penguin Press.

American National Election Studies. 2016. "2016 Time Series Study." https://electionstudies.org/data-center/2016-time-series-study/.

►**Anderson, Michael L.** 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association* 103 (484): 1481–95.

►**Appel, Helmut, Alexander L. Gerlach, and Jan Crusius.** 2016. "The Interplay between Facebook Use, Social Comparison, Envy, and Depression." *Current Opinion in Psychology* 9: 44–49.

Argyle, Michael. 2001. *The Psychology of Happiness.* London: Routledge.

►**Athey, Susan, Julie Tibshirani, and Stefan Wagner.** 2019. "Generalized Random Forests." *Annals of Statistics* 47 (2): 1148–78.

Baker, David A., and Guillermo Perez Algorta. 2016. "The Relationship between Online Social Networking and Depression: A Systematic Review of Quantitative Studies." *Cyberpsychology, Behavior, and Social Networking* 19 (11).

►**Bakshy, Eytan, Solomon Messing, and Lada A. Adamic.** 2015. "Exposure to Ideologically Diverse News and Opinion on Facebook." *Science* 348 (6239): 1130–32.

►**Bartels, Larry M.** 1993. "Messages Received: The Political Impact of Media Exposure." *American Political Science Review* 87 (2): 267–85.

Baym, Nancy K., Kelly B. Wagman, and Christopher J. Persaud. Forthcoming. "Mindfully Scrolling: Rethinking Facebook after Time Deactivated." *Social Media + Society.*

Becker, Gary S., Michael Grossman, and Kevin M. Murphy. 1991. "Rational Addiction and the Effect of Price on Consumption." *American Economic Review* 81 (2): 237–41.

►**Becker, Gary S., and Kevin M. Murphy.** 1988. "A Theory of Rational Addiction." *Journal of Political Economy* 96 (4): 675–700.

►**Becker, Gordon M., Morris H. Degroot, and Jacob Marschak.** 1964. "Measuring Utility by a Single-Response Sequential Method." *Behavioral Science* 9 (3): 226–32.

►**Benjamin, Daniel J., Ori Heffetz, Miles S. Kimball, and Alex Rees-Jones.** 2012. "What Do You Think Would Make You Happier? What Do You Think You Would Choose?" *American Economic Review* 102 (5): 2083–2110.

►**Benjamini, Yoav, Abba M. Krieger, and Daniel Yekutieli.** 2006. "Adaptive Linear Step-Up Procedures that Control the False Discovery Rate." *Biometrika* 93 (3): 491–507.

►**Besley, Timothy, and Robin Burgess.** 2001. "Political Agency, Government Responsiveness and the Role of the Media." *European Economic Review* 45 (4–6): 629–40.

►**Bohm, Peter, Johan Lindén, and Joakim Sonnegård.** 1997. "Eliciting Reservation Prices: Becker-DeGroot-Marschak Mechanisms vs. Markets." *Economic Journal* 107 (443): 1079–89.

►**Bolier, Linda, Merel Haverman, Gerben J. Westerhof, Heleen Riper, Filip Smit, and Ernst Bohlmeijer.** 2013. "Positive Psychology Interventions: A Meta-Analysis of Randomized Controlled Studies." *BMC Public Health* 13: 119.

Boxell, Levi. 2018. "Demographic Change and Political Polarization in the United States." https://ssrn.com/abstract=3148805.

Brynjolfsson, Erik, Felix Eggers, and Avinash Gannamaneni. 2018. "Using Massive Online Choice Experiments to Measure Changes in Well-Being." NBER Working Paper 24514.

Brynjolfsson, Erik, and Joo Hee Oh. 2012. "The Attention Economy: Measuring the Value of Free Digital Services on the Internet." Unpublished.

Brynjolfsson, Erik, Daniel Rock, and Chad Syverson. 2019. "Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics." In *The Economics of Artificial Intelligence: An Agenda,* edited by Ajay Agrawal, Joshua Gans, and Avi Goldfarb, 23–60. Chicago: University of Chicago Press.

Brynjolfsson, Erik, and Adam Saunders. 2009. "What the GDP Gets Wrong (Why Managers Should Care)." *MIT Sloan Management Review* 51 (1): 95.

Burke, Moira, and Robert E. Kraut. 2014. "Growing Closer on Facebook: Changes in Tie Strength through Social Network Site Use." *CHI '14: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems:* 4187–96.

►**Burke, Moira, and Robert E. Kraut.** 2016. "The Relationship between Facebook Use and Well-Being Depends on Communication Type and Tie Strength." *Journal of Computer-Mediated Communication* 21 (4): 265–81.

Burke, Moira, Robert E. Kraut, and Cameron Marlow. 2011. "Social Capital on Facebook: Differentiating Uses and Users." *CHI '11: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems:* 571–80.

Burke, Moira, Cameron Marlow, and Thomas Lento. 2010. "Social Network Activity and Social Well-Being." *CHI '10: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*: 1909–12.

Busse, Meghan R., Devin G. Pope, Jaren C. Pope, and Jorge Silva-Risso. 2015. "The Psychological Effect of Weather on Car Purchases." *Quarterly Journal of Economics* 130 (1): 371–414.

Byrne, David M., John G. Fernald, and Marshall B. Reinsdorf. 2016. "Does the United States Have a Productivity Slowdown or a Measurement Problem?" *Brookings Papers on Economic Activity* (Spring): 109–82.

Charness, Gary, and Uri Gneezy. 2009. "Incentives to Exercise." *Econometrica* 77 (3): 909–31.

Chopik, William J. 2017. "Associations among Relational Values, Support, Health, and Well-Being across the Adult Lifespan." *Personal Relationships* 24 (2): 408–22.

Conlin, Michael, Ted O'Donoghue, and Timothy J. Vogelsang. 2007. "Projection Bias in Catalog Orders." *American Economic Review* 97 (4): 1217–49.

Corrigan, Jay R., Saleem Alhabash, Matthew Rousu, and Sean B. Cash. 2018. "How Much Is Social Media Worth? Estimating the Value of Facebook by Paying Users to Stop Using It." *PLOS ONE* 13 (12): e0207101.

Csikszentmihalyi, Mihaly, and Reed Larson. 2014. "Validity and Reliability of the Experience-Sampling Method." In *Flow and the Foundations of Positive Psychology*, 35–54. New York: Springer.

DellaVigna, Stefano, and Matthew Gentzkow. 2010. "Persuasion: Empirical Evidence." *Annual Review of Economics* 2: 643–69.

DellaVigna, Stefano, and Ethan Kaplan. 2007. "The Fox News Effect: Media Bias and Voting." *Quarterly Journal of Economics* 122 (3): 1187–1234.

DellaVigna, Stefano, and Eliana La Ferrara. 2015. "Economic and Social Impacts of the Media." In *Handbook of Media Economics*, Vol. 1A, edited by Simon P. Anderson, David Strömberg, and Joel Waldfogel, 723–68. Amsterdam: Elsevier.

Deters, Fenne Große, and Matthias R. Mehl. 2013. "Does Posting Facebook Status Updates Increase or Decrease Loneliness? An Online Social Networking Experiment." *Social Psychological and Personality Science* 4 (5): 579–86.

Diener, Ed, Robert A. Emmons, Randy J. Larsen, and Sharon Griffin. 1985. "The Satisfaction with Life Scale." *Journal of Personality Assessment* 49 (1): 71–75.

Ellison, Nicole B., Charles Steinfield, and Cliff Lampe. 2007. "The Benefits of Facebook 'Friends': Social Capital and College Students' Use of Online Social Network Sites." *Journal of Computer-Mediated Communication* 12 (4): 1143–68.

Enikolopov, Ruben, Alexey Makarin, and Maria Petrova. 2018. "Social Media and Protest Participation: Evidence from Russia." https://ssrn.com/abstract=2696236.

Enikolopov, Ruben, and Maria Petrova. 2015. "Media Capture: Empirical Evidence." In *Handbook of Media Economics*, Vol. 1A, edited by Simon P. Anderson, David Strömberg, and Joel Waldfogel, 687–700. Amsterdam: Elsevier.

Enikolopov, Ruben, Maria Petrova, and Ekaterina Zhuravskaya. 2011. "Media and Political Persuasion: Evidence from Russia." *American Economic Review* 101 (7): 3253–85.

Facebook. 2016. "Facebook Q1 2016 Results: Earnings Call Transcript." *Seeking Alpha*, April 27, 2016. https://seekingalpha.com/article/3968783-facebook-fb-mark-elliot-zuckerberg-q1-2016-results-earnings-call-transcript.

Facebook. 2018. "Facebook Reports Third Quarter 2018 Results." Press Release. https://investor.fb.com/investor-news/press-release-details/2018/Facebook-Reports-Third-Quarter-2018-Results/default.aspx.

Fiorina, Morris P., and Samuel J. Abrams. 2008. "Political Polarization in the American Public." *Annual Review Political Science* 11: 563–88.

Frison, Eline, and Steven Eggermont. 2015. "Toward an Integrated and Differential Approach to the Relationships between Loneliness, Different Types of Facebook Use, and Adolescents' Depressed Mood." *Communication Research*. doi.org/10.1177/0093650215617506.

Fujiwara, Thomas, Kyle Meng, and Tom Vogl. 2016. "Habit Formation in Voting: Evidence from Rainy Elections." *American Economic Journal: Applied Economics* 8 (4): 160–88.

Gentzkow, Matthew. 2006. "Television and Voter Turnout." *Quarterly Journal of Economics* 121 (3): 931–72.

Gentzkow, Matthew. 2016. "Polarization in 2016." Unpublished.

Gentzkow, Matthew, and Jesse M. Shapiro. 2011. "Ideological Segregation Online and Offline." *Quarterly Journal of Economics* 126 (4): 1799–1839.

►**Gerber, Alan S., James G. Gimpel, Donald P. Green, and Daron R. Shaw.** 2011. "How Large and Long-Lasting Are the Persuasive Effects of Televised Campaign Ads? Results from a Randomized Field Experiment." *American Political Science Review* 105 (1): 135–50.
►**Gerber, Alan S., and Donald P. Green.** 2000. "The Effects of Canvassing, Telephone Calls, and Direct Mail on Voter Turnout: A Field Experiment." *American Political Science Review* 94 (3): 653–63.
►**Gerber, Alan S., Dean Karlan, and Daniel Bergan.** 2009. "Does the Media Matter? A Field Experiment Measuring the Effect of Newspapers on Voting Behavior and Political Opinions." *American Economic Journal: Applied Economics* 1 (2): 35–52.
►**Gerson, Jennifer, Anke C. Plagnol, and Philip J. Corr.** 2017. "Passive and Active Facebook Use Measure (PAUM): Validation and Relationship to the Reinforcement Sensitivity Theory." *Personality and Individual Differences* 117: 81–90.
►**Gruber, Jonathan, and Botond Kőszegi.** 2001. "Is Addiction 'Rational'? Theory and Evidence." *Quarterly Journal of Economics* 116 (4): 1261–1303.
Howard, Philip N., Aiden Duffy, Deen Freelon, Muzammil M. Hussain, Will Mari, and Marwa Mazaid. 2011. "Opening Closed Regimes: What Was the Role of Social Media During the Arab Spring?" https://ssrn.com/abstract=2595096.
►**Huber, Gregory A., and Kevin Arceneaux.** 2007. "Identifying the Persuasive Effects of Presidential Advertising." *American Journal of Political Science* 51 (4): 957–77.
►**Hughes, Mary Elizabeth, Linda J. Waite, Louise C. Hawkley, and John T. Cacioppo.** 2004. "A Short Scale for Measuring Loneliness in Large Surveys: Results from Two Population-Based Studies." *Research on Aging* 26 (6): 655–72.
►**Huppert, Felicia A., Nic Marks, Andrew Clark, Johannes Siegrist, Alois Stutzer, Joar Vittersø, and Morten Wahrendorf.** 2009. "Measuring Well-Being across Europe: Description of the ESS Well-Being Module and Preliminary Findings." *Social Indicators Research* 91 (3): 301–15.
Hussam, Reshmaan, Atonu Rabbani, Giovanni Reggiani, and Natalia Rigol. 2016. "Handwashing and Habit Formation." Unpublished.
►**Iyengar, Shanto, Gaurav Sood, and Yphtach Lelkes.** 2012. "Affect, Not Ideology: Social Identity Perspective on Polarization." *Public Opinion Quarterly* 76 (3): 405–31.
►**Iyengar, Shanto, and Sean J. Westwood.** 2015. "Fear and Loathing across Party Lines: New Evidence on Group Polarization." *American Journal of Political Science* 59 (3): 690–707.
►**Kahneman, Daniel, Alan B. Krueger, David Schkade, Norbert Schwarz, and Arthur A. Stone.** 2006. "Would You Be Happier If You Were Richer? A Focusing Illusion." *Science* 312 (5782): 1908–10.
Kirkpatrick, David. 2011. *The Facebook Effect: The Inside Story of the Company That Is Connecting the World.* New York: Simon & Schuster.
Krasnova, Hanna, Helena Wenninger, Thomas Widjaja, and Peter Buxmann. 2013. "Envy on Facebook: A Hidden Threat to Users' Life Satisfaction." Unpublished.
►**Kross, Ethan, Philippe Verduyn, Emre Demiralp, Jiyoung Park, David Seungjae Lee, Natalie Lin, Holly Shablack, John Jonides, and Oscar Ybarra.** 2013. "Facebook Use Predicts Declines in Subjective Well-Being in Young Adults." *PLOS ONE* 8 (8): e69841.
►**Lee, David S.** 2009. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." *Review of Economic Studies* 76 (3): 1071–1102.
►**Loewenstein, George, Ted O'Donoghue, and Matthew Rabin.** 2003. "Projection Bias in Predicting Future Utility." *Quarterly Journal of Economics* 118 (4): 1209–48.
►**Lyubomirsky, Sonja, and Heidi S. Lepper.** 1999. "A Measure of Subjective Happiness: Preliminary Reliability and Construct Validation." *Social Indicators Research* 46 (2): 137–55.
►**Martin, Gregory J., and Ali Yurukoglu.** 2017. "Bias in Cable News: Persuasion and Polarization." *American Economic Review* 107 (9): 2565–99.
►**Mazar, Nina, Botond Kőszegi, and Dan Ariely.** 2014. "True Context-Dependent Preferences? The Causes of Market-Dependent Valuations." *Journal of Behavioral Decision Making* 27 (3): 200–208.
Molla, Rani, and Kurt Wagner. 2018. "People Spend Almost as Much Time on Instagram as They Do on Facebook." *Recode*, June 25, 2018. https://www.recode.net/2018/6/25/17501224/instagram-facebook-snapchat-time-spent-growth-data.
Mosquera, Roberto, Mofioluwasademi Odunowo, Trent McNamara, Xiongfei Guo, and Ragan Petrie. 2018. "The Economic Effects of Facebook." https://ssrn.com/abstract=3312462.
Müller, Karsten, and Carlo Schwarz. 2018. "Fanning the Flames of Hate: Social Media and Hate Crime." https://ssrn.com/abstract=3082972.
►**Myers, David G.** 2000. "The Funds, Friends, and Faith of Happy People." *American Psychologist* 55 (1): 56–67.
Nakamura, Leonard I., Jon Samuels, and Rachel H. Soloveichik. 2016. "Valuing 'Free' Media in GDP: An Experimental Approach." Federal Reserve Bank of Philadelphia Working Paper 16-24.

Napoli, Philip M. 2014. "Measuring Media Impact." The Norman Lear Center. https://learcenter.org/pdf/measuringmedia.pdf.

Newport, Cal. 2019. *Digital Minimalism: Choosing a Focused Life in a Noisy World.* New York: Penguin.

▶ **Olken, Benjamin A.** 2009. "Do Television and Radio Destroy Social Capital? Evidence from Indonesian Villages." *American Economic Journal: Applied Economics* 1 (4): 1–33.

Pew Research Center. 2018. "News Use across Social Media Platforms 2018." http://www.journalism.org/2018/09/10/news-use-across-social-media-platforms-2018/.

▶ **Reis, Harry T., W. Andrew Collins, and Ellen Berscheid.** 2000. "The Relationship Context of Human Behavior and Development." *Psychological Bulletin* 126 (6): 844–72.

▶ **Sagioglu, Christina, and Tobias Greitemeyer.** 2014. "Facebook's Emotional Consequences: Why Facebook Causes a Decrease in Mood and Why People Still Use It." *Computers in Human Behavior* 35: 359–63.

▶ **Satici, Seydi Ahmet, and Recep Uysal.** 2015. "Well-Being and Problematic Facebook Use." *Computers in Human Behavior* 49: 185–90.

Settle, Jaime E. 2018. *Frenemies: How Social Media Polarizes America.* New York: Cambridge University Press.

▶ **Shakya, Holly B., and Nicholas A. Christakis.** 2017. "Association of Facebook Use with Compromised Well-Being: A Longitudinal Study." *American Journal of Epidemiology* 185 (3): 203–11.

▶ **Simonsohn, Uri.** 2010. "Weather to Go to College." *Economic Journal* 120 (543): 270–80.

Spenkuch, Jörg L., and David Toniatti. 2016. "Political Advertising and Election Outcomes." CESifo Working Paper 5780.

▶ **Stone, Arthur A., and Saul Shiffman.** 1994. "Ecological Momentary Assessment (EMA) in Behavioral Medicine." *Annals of Behavioral Medicine* 16 (3): 199–202.

▶ **Strömberg, David.** 2015. "Media and Politics." *Annual Review of Economics* 7: 173–205.

Sunstein, Cass R. 2001. *Republic.com.* Princeton, NJ: Princeton University Press.

Sunstein, Cass R. 2017. *#Republic: Divided Democracy in the Age of Social Media.* Princeton, NJ: Princeton University Press.

Sunstein, Cass R. Forthcoming. "Valuing Facebook." *Behavioural Public Policy.*

▶ **Syverson, Chad.** 2017. "Challenges to Mismeasurement Explanations for the US Productivity Slowdown." *Journal of Economic Perspectives* 31 (2): 165–86.

▶ **Tandoc, Edson C., Patrick Ferrucci, and Margaret Duffy.** 2015. "Facebook Use, Envy, and Depression among College Students: Is Facebooking Depressing?" *Computers in Human Behavior* 43: 139–46.

Twenge, Jean M. 2017. *iGen: Why Today's Super-Connected Kids Are Growing Up Less Rebellious, More Tolerant, Less Happy—and Completely Unprepared for Adulthood—and What That Means for the Rest of Us.* New York: Simon & Schuster.

▶ **Twenge, Jean M., Thomas E. Joiner, Megan L. Rogers, and Gabrielle N. Martin.** 2018. "Increases in Depressive Symptoms, Suicide-Related Outcomes, and Suicide Rates among U.S. Adolescents after 2010 and Links to Increased New Media Screen Time." *Clinical Psychological Science* 6 (1): 3–17.

▶ **Twenge, Jean M., Gabrielle N. Martin, and W. Keith Campbell.** 2018. "Decreases in Psychological Well-Being among American Adolescents after 2012 and Links to Screen Time during the Rise of Smartphone Technology." *Emotion* 18 (6): 765–80.

Twenge, Jean M., and Heejung Park. 2019. "The Decline in Adult Activities among U.S. Adolescents, 1976–2016." *Child Development* 90 (2): 638–54.

Twenge, Jean M., Ryne A. Sherman, and Sonja Lyubomirsky. 2016. "More Happiness for Young People and Less for Mature Adults: Time Period Differences in Subjective Well-Being in the United States, 1972–2014." *Social Psychological and Personality Science* 7 (2): 131–41.

US Census Bureau. 2017. "2017 American Community Survey 1-Year Estimates." https://www.census.gov/newsroom/press-kits/2018/acs-1year.html.

▶ **Vanden Abeele, Mariek M. P., Marjolijn L. Antheunis, Monique M. H. Pollmann, Alexander P. Schouten, Christine C. Liebrecht, Per J. van der Wijst, Marije A. A. van Amelsvoort et al.** 2018. "Does Facebook Use Predict College Students' Social Capital? A Replication of Ellison, Steinfield, and Lampe's (2007) Study Using the Original and More Recent Measures of Facebook Use and Social Capital." *Communication Studies* 69 (3): 272–82.

▶ **Verduyn, Philippe, David Seungjae Lee, Jiyoung Park, Holly Shablack, Ariana Orvell, Joseph Bayer, Oscar Ybarra, John Jonides, and Ethan Kross.** 2015. "Passive Facebook Usage Undermines Affective Well-Being: Experimental and Longitudinal Evidence." *Journal of Experimental Psychology* 144 (2): 480–88.

This article has been cited by:

1. Hunt Allcott, Matthew Gentzkow, Lena Song. 2022. Digital Addiction. *American Economic Review* **112**:7, 2424-2463. [Abstract] [View PDF article] [PDF with links]
2. Ro'ee Levy. 2021. Social Media, News Consumption, and Polarization: Evidence from a Field Experiment. *American Economic Review* **111**:3, 831-870. [Abstract] [View PDF article] [PDF with links]

NBER WORKING PAPER SERIES

THE EFFECT OF DEACTIVATING FACEBOOK AND INSTAGRAM ON USERS' EMOTIONAL STATE

Hunt Allcott
Matthew Gentzkow
Benjamin Wittenbrink
Juan Carlos Cisneros
Adriana Crespo-Tenorio
Drew Dimmery
Deen Freelon
Sandra González-Bailón
Andrew M. Guess
Young Mie Kim
David Lazer
Neil Malhotra
Devra Moehler
Sameer Nair-Desai
Brendan Nyhan
Jennifer Pan
Jaime Settle
Emily Thorson
Rebekah Tromble
Carlos Velasco Rivera
Arjun Wilkins
Magdalena Wojcieszak
Annie Franco
Chad Kiewiet de Jonge
Winter Mason
Natalie Jomini Stroud
Joshua A. Tucker

Experiment

The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w33697

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

The Effect of Deactivating Facebook and Instagram on Users' Emotional State
Hunt Allcott, Matthew Gentzkow, Benjamin Wittenbrink, Juan Carlos Cisneros, Adriana Crespo-Tenorio, Drew Dimmery, Deen Freelon, Sandra González-Bailón, Andrew M. Guess, Young Mie Kim, David Lazer, Neil Malhotra, Devra Moehler, Sameer Nair-Desai, Brendan Nyhan, Jennifer Pan, Jaime Settle, Emily Thorson, Rebekah Tromble, Carlos Velasco Rivera, Arjun Wilkins, Magdalena Wojcieszak, Annie Franco, Chad Kiewiet de Jonge, Winter Mason, Natalie Jomini Stroud, and Joshua A. Tucker
NBER Working Paper No. 33697
April 2025
JEL No. I1, L82

ABSTRACT

We estimate the effect of social media deactivation on users' emotional state in two large randomized experiments before the 2020 U.S. election. People who deactivated Facebook for the six weeks before the election reported a 0.060 standard deviation improvement in an index of happiness, depression, and anxiety, relative to controls who deactivated for just the first of those six weeks. People who deactivated Instagram for those six weeks reported a 0.041 standard deviation improvement relative to controls. Exploratory analysis suggests the Facebook effect is driven by people over 35, while the Instagram effect is driven by women under 25.

Hunt Allcott
Stanford University
and NBER
allcott@stanford.edu

Matthew Gentzkow
Stanford University
Department of Economics
and NBER
gentzkow@stanford.edu

Benjamin Wittenbrink
Massachusetts Institute of Technology
bwitten@mit.edu

Juan Carlos Cisneros
Universitat Pompeu Fabra
Department of Economics and Business
juan.cisneros@bse.eu

Adriana Crespo-Tenorio
Meta
adrianact@gmail.com

Drew Dimmery
Meta
drew.dimmery@gmail.com

Deen Freelon
University of Pennsylvania
deen.freelon@asc.upenn.edu

Sandra González-Bailón
University of Pennsylvania
sandra.gonzalez.bailon@asc.upenn.edu

Andrew M. Guess
Princeton University
School of Public and International Affairs
Department of Politics
aguess@princeton.edu

Young Mie Kim
University of Wisconsin - Madison
ymkim5@wisc.edu

David Lazer
Northeastern University
Khoury College of Computer Sciences
d.lazer@northeastern.edu

Neil Malhotra
Stanford University
neilm@stanford.edu

Devra Moehler
Meta
dmoehler@fb.com

Sameer Nair-Desai
University of Michigan, Ann Arbor
nairdesa@umich.edu

Brendan Nyhan
Dartmouth College
nyhan@dartmouth.edu

Jennifer Pan
Stanford University
jp1@stanford.edu

Jaime Settle
College of William & Mary
jsettle@wm.edu

Emily Thorson
Syracuse University
eathorso@syr.edu

Rebekah Tromble
The George Washington University
rtromble@email.gwu.edu

Carlos Velasco Rivera
Meta
carlos.velascorivera@gmail.com

Arjun Wilkins
Meta
awilkins@meta.com

Magdalena Wojcieszak
University of California, Davis
Department of Communications and
University of Warsaw, Center
for Excellence in Social Science
mwojcieszak@ucdavis.edu

Annie Franco
Meta
afranco@meta.com

Chad Kiewiet de Jonge
Meta
chadkdj@meta.com

Winter Mason
Meta
m@winteram.com

Natalie Jomini Stroud University of
Texas at Austin
tstroud@austin.utexas.edu

Joshua A. Tucker New York
University joshua.tucker@nyu.edu

A data appendix is available at: http://www.nber.org/data-appendix/w33697
A randomized controlled trials registry entry is available at: https://osf.io/t9q2f

There is an active debate over how social media affect users' psychological well-being. Do social media make people happier, for example by facilitating beneficial social connections?[1] Or do they make people depressed and anxious, for example by reducing face-to-face interactions or increasing unfavorable social comparisons?[2] Some analysts argue that social media have contributed to the alarming recent decline in young people's mental health,[3] and policymakers have responded with legislation and legal action.[4] These high-stakes debates have relied primarily on evidence from time-series and cross-sectional correlations, plus a few relatively small randomized experiments, and scholars disagree on the implications (Odgers 2024; Dubner 2024; Capraro et al. 2024).

In a separate trend, American elections have become increasingly stressful: an August 2020 study found that 68 percent of American adults cited the upcoming election as a significant source of stress, a material increase from 2016 (American Psychological Association 2020). Other studies find that exposure to political news reduces psychological well-being (Pierce, Rogers and Snyder 2016; Simchon et al. 2020; Gray, Pickard and Munford 2021; Ford et al. 2023; Kimball et al. 2024). Since many people get political news on social media (Allcott and Gentzkow 2017), these facts raise the question of how using social media before an election affects people's emotional state.

In this paper, we report the results of the largest-ever experimental study on the effect of social media deactivation on users' emotional state, which we carried out as part of a broader study of political outcomes before the 2020 U.S. presidential election. We recruited 19,857

[1] See Reis, Collins and Berscheid (2000), Chopik (2017), and Chetty et al. (2022) on the importance of human connection and social capital.

[2] See Verduyn et al. (2015), Tromholt (2016), Hunt et al. (2018), Turel, Cavagnaro and Meshi (2018), Cohen et al. (2019), Brailovskaia et al. (2020), Mosquera et al. (2020), Ozimek and Bierhoff (2020), Siegel (2020), Castaño-Pulgarín et al. (2021), Przybylski et al. (2021), van Wezel et al. (2021), Collis and Eggers (2022), Brailovskaia et al. (2023), González-Bailón et al. (2023), and Thai et al. (2023).

[3] See Twenge (2017), Engeln et al. (2020), Office of the Surgeon General (2021), Wells, Horwitz and Seetharaman (2021), and Haidt (2023).

[4] See Archie (2023), New York State Attorney General (2023), Utah Governor's Office (2023), European Commission (2024), and National Conference of State Legislatures (2024).

Facebook users and 15,585 Instagram users who spent at least 15 minutes per day on the respective platform. We randomly assigned 27 percent of participants to a treatment group that was offered payment for deactivating their accounts for the six weeks before the election. The remaining participants formed a control group that was paid to deactivate for just the first of those six weeks. Our baseline and endline surveys elicited three measures of self-reported emotional state—how much of the time during the past four weeks that people felt happy, depressed, or anxious—along with a large suite of political outcomes that we study separately in Allcott et al. (2024).

We estimate that users in the Facebook deactivation group reported a 0.060 standard deviation improvement in an index of happiness, anxiety, and depression, relative to control users. The effect is statistically distinguishable from zero at the $p < 0.01$ level, both when considered individually and after adjusting for multiple hypothesis testing along with the full set of political outcomes considered in Allcott et al. (2024). Non-preregistered subgroup analyses suggest larger effects of Facebook on people over 35, undecided voters, and people without a college degree.

We estimate that users in the Instagram deactivation group reported a 0.041 standard deviation improvement in the emotional state index relative to control. The effect is statistically distinguishable from zero at the $p = 0.016$ level when considered individually, and at the $p = 0.14$ level after adjusting for multiple hypothesis testing along with the outcomes in Allcott et al. (2024). The latter estimate does not meet our pre-registered $p = 0.05$ significance threshold. Substitution analyses imply this improvement is achieved without shifts to offline activities. Non-preregistered subgroup analyses suggest larger effects of Instagram on women aged 18-24.

We offer several points of comparison for the effect sizes. After controlling for other demographics, emotional state is 0.48 standard deviations higher for Republicans than for Democrats

in our sample. The average psychological intervention reported in the van Agteren et al. (2021) meta-analysis improved emotional state by 0.27 standard deviations. Finally, a different index of young people's emotional state worsened by 0.37 standard deviations between 2008 and 2022.

Two additional results provide insight into mechanisms. First, app use data show that when people deactivate, most of time freed by Facebook deactivation and all of time freed by Instagram deactivation is substituted to other smartphone apps. Thus, we expect no effect on offline time for individuals in the Instagram group and some moderate effects on offline time for those in the Facebook group. Instead, this suggests that the effects are mostly driven by the different user experiences of Facebook or Instagram compared to other apps. Second, the effects are not significantly different for people with higher online or offline baseline political engagement. This provides no evidence that the effects are driven by factors specific to the election period.

Our work relates to a large literature on the relationship between social media use and emotional state. The vast majority of prior studies focus on non-experimental results such as cross-sectional or longitudinal correlations; the recent Hancock et al. (2022) meta-analysis included 226 correlation studies. We find that in our data, standard non-experimental approaches yield results that are biased in unpredictable directions relative to the experimental estimates.

We are aware of seven prior experiments estimating the effects of at least one week of social media abstention.[3] The largest prior experiment, Allcott et al. (2020), was written by two of the lead authors of this study, and we reused key elements of that earlier design. Our work goes significantly beyond Allcott et al. (2020) and the other prior experiments in several ways. First, our experiment is the first to specifically estimate the impact of Instagram access. Given the different user experiences and populations on Instagram versus Facebook, it would have been unclear whether effects of Facebook deactivation translate to Instagram. The second is

[3]The seven experiments are Tromholt (2016), Turel et al. (2018), Allcott et al. (2020), Mosquera et al. (2020), Hall et al. (2021), Lambert et al. (2022), and Arceneaux et al. (2023); see Appendix Table S30.

size and scope: our total sample is about 20 times larger than in the largest prior experiment, and we required longer abstention than any prior study. Given this much larger sample size, we are better powered for subgroup analysis; the heterogeneous effects we document by age and gender had gone undetected in previous experimental work. The third is data: we are the only experiment to leverage internal data from a social media platform and to meter substitution to other smartphone apps. Our result that most of time freed by deactivation is substituted to other apps is crucial to understanding mechanisms, but Allcott et al. (2020) and others had relied on self-reports that did not make this clear. Fourth, ours is the only experiment to be carried out in the context of a U.S. presidential election. Social media use before a presidential election could have different effects on emotional state than use before midterm elections in non-election periods, given heightened media coverage and perceived higher stakes.

Allcott et al. (2024) studies the same experiment and is covered by the same pre-analysis plan, but that paper considers only political outcomes, not emotional state. We ran this study during the election period because our primary goal was to estimate effects on political outcomes. Our work also builds on quasi-experimental evidence in Braghieri, Levy and Makarin (2022), which finds that Facebook worsened college students' mental health when it was rolled out in 2004 and 2005. Facebook's user experience was very different during that rollout period—for example, there was no news feed—so the effects could be quite different two decades later.

We also build on prior work on emotional responses to elections. Existing studies primarily describe the time path of emotional state for different groups, including the response to election outcomes (e.g., Pierce et al. 2016; Suzuki et al. 2023; Kimball et al. 2024). We focus on how social media use during an election season affects emotional state.

This paper has several important limitations. First, our findings are only directly informative about the people who agreed to participate and deactivate their accounts for the payments we offered. While we use weights to adjust for sample selection, our sample may also be selected

on unobserved characteristics. Second, our experiment measures the effects of an incremental five weeks of individual deactivation before the 2020 election. Effects could differ for longer-term deactivation, simultaneous deactivation of many users, deactivation during a non-election period, or deactivation during a future election. For context, about six percent of Facebook content viewed in 2020 was politics-related (Schultz 2020), and Facebook and Instagram have both reduced political content in news feeds since 2020 (Meta 2024). Third, our analysis relies on three specific self-reported emotional state survey questions in the context of a larger survey focused on political outcomes. Effects could differ for other emotional state measures or on a different survey instrument (see, e.g., Zaller and Feldman (1992)). Fourth, although we designed the experiment to mitigate experimenter demand effects and previous work suggests that such effects may be limited in our context (De Quidt, Haushofer and Roth 2018; Mummolo and Peterson 2019; Allcott et al. 2020), we cannot definitively rule out the possibility that survey responses were influenced by participants' knowledge that they were in an experiment. Finally, the baseline emotional state outcomes were imbalanced by chance in the Facebook sample, and not all participants completed the endline surveys, although the data suggest that these issues are unlikely to substantially affect the results.

This project is part of the U.S. 2020 Facebook and Instagram Election Study (González-Bailón et al. 2023; Guess et al. 2023a,b; Nyhan et al. 2023; Allcott et al. 2024), a partnership between Meta researchers and unpaid independent academics. Under the terms of the collaboration, the independent academic authors had final authority over the pre-analysis plan, data analysis, and manuscript text, and Meta could not block any results from being published. More details of this partnership are in Appendices E and F. We have posted answers to frequently asked questions online at this website.

Sections 1-5, respectively, present the experimental design, descriptive statistics, empirical strategy, impact evaluation results, and conclusion.

1 Experimental Design

We ran two parallel experiments, with Facebook and Instagram as the respective "focal platform." For each focal platform, Meta drew a stratified random sample of users who were in the U.S., were age 18 or older, and had logged in at least once in the past month. From August 31 to September 12, Meta placed survey invitations at the top of these users' focal platform news feeds. People who clicked on the invitations were told about the study and asked what weekly payments they would be willing to accept to deactivate their focal platform accounts for either one or six weeks. Those who were willing to deactivate for $25 per week and consented to participate were immediately sent to the National Opinion Research Corporation (NORC) website to complete a short enrollment survey. NORC fielded the baseline and endline surveys on September 8-21 and November 4-18, respectively.

Participants who completed the baseline survey were randomized into two groups: Deactivation (27 percent) and Control (73) percent.[6] The Deactivation group was told that they would receive $150 if they did not log into the focal platform for the next six weeks, while the Control group was told that they would receive $25 if they did not log in for the next week.[7] By including a short deactivation period for the Control group, we guaranteed that the only differences between Control and Deactivation were the total payment amount and deactivation length, and that all participants would perceive themselves to be part of a study that involved deactivating their social media accounts. This reduced the risk of experimenter demand effects, differential attrition, and any spurious effects that might be artifacts of the deactivation process itself.

Meta deactivated participants' focal platform accounts starting September 23. Control and Deactivation group accounts were automatically reactivated on September 30 and November 4,

[6]Randomization was stratified into 36 strata defined by an indicator for residence in an election swing state, average daily time spent on Facebook or Instagram over the previous 30 days, self-reported political party identification, and race.

[7]Since the Deactivation condition was more expensive than Control, we allocated a smaller share to the former to increase power per dollar of cost.

respectively.

The research goal was to evaluate the core Facebook and Instagram products. Thus, we allowed participants to still use their Facebook and Instagram credentials to access other apps, including WhatsApp and Facebook Messenger. Participants could log back into the Facebook or Instagram app and reactivate their accounts at any time. People who logged in were reminded that they would lose their deactivation payments but were asked to complete the remaining surveys.

Participants were paid at least $5 for completing the baseline survey and at least $20 for endline. Participants were also offered additional payment to allow passive tracking of their smartphone app and web browser use.

1.1 Emotional State Survey Questions

The baseline and endline surveys covered demographics and a wide range of political beliefs and behaviors. This paper focuses on the three emotional state questions that were also included: "Please tell us how much of the time during the past four weeks you felt [happy / depressed / anxious]." The response options were "All of the time," "Often," "Sometimes," "Rarely," and "Never." We code these responses as 4, 3, 2, 1, and 0, respectively, and then standardize each to have a mean of zero and a standard deviation of one in the sample-weighted combined Control groups, giving the variables *happy*, *depressed*, and *anxious*. We re-sign *depressed* and *anxious* as *depressed x (-1)* and *anxious x (-1)*, so that more positive values correspond to more positive emotional state. The *emotional state index* is the average of these three re-signed variables, re-standardized to have a standard deviation of one in the sample-weighted combined Control groups.

These three questions were drawn from the European Social Survey Well-being Module (Huppert et al. 2009) and are similar to other established emotional state measures. The *emo-*

tional state index has a Cronbach's alpha of 0.79 and 0.78 in the Facebook and Instagram samples, respectively, and is correlated in expected ways with demographics and other emotional state measures. See Appendix B for more information on item development, reliability, and validity.

1.2 Pre-Analysis Plan, Multiple Hypothesis Testing, and Subgroup Analysis

We submitted an initial pre-analysis plan (PAP) on September 22 and a slightly updated final PAP on November 3rd, the day before the endline survey began. We followed the PAP in all respects, with the exception of six minor deviations or clarifications reported in Appendix G, which were mainly driven by changes in data availability relative to what we anticipated. All of the pre-specified analyses except for effects on emotional state outcomes are presented in Allcott et al. (2024).

The PAP originally implied that results for all outcomes would be presented in a single paper. However, as we drafted the paper, it became clear that it was not possible to fully present the motivation, related literature, robustness, and interpretation for both the political and emotional state outcomes in a single paper, so we split the results into two.

The PAP stated that the primary analysis sample would include only participants with greater than 15 minutes per day baseline use, as we expected that deactivation would have limited effects for people who rarely use the platform.

The PAP stated that the *emotional state index* was a "secondary outcome," and the three individual components of emotional state were "auxiliary outcomes." To control for multiple hypothesis testing, the PAP stated that for *emotional state index* and all other secondary outcomes, we would present sharpened False Discovery Rate (FDR) adjusted q-values (Benjamini, Krieger and Yekutieli 2006) adjusted against all 61 primary and secondary outcomes, including

political beliefs, political polarization, and voting behavior. We also present unadjusted p-values on *emotional state index*, which may be relevant to researchers with a specific a priori interest in effects on emotional state (Kling, Liebman and Katz 2007; Casey, Glennerster and Miguel 2012). The PAP stated that we would report unadjusted p-values on auxiliary outcomes such as the three emotional state components. Wasserstein and Lazar (2016) and Imbens (2021) argue against overemphasizing binary statements about whether a result is "statistically significant." The PAP stated that for any such statements we do make, we would use two-sided tests with a $p < 0.05$ threshold.

The pre-analysis plan described how we would carry out subgroup analyses for the experiment's primary outcomes, but it did not specify what subgroup analysis might be done for secondary outcomes such as emotional state. The preregistered subgroups were chosen with the primary political outcomes in mind, and some are thus less relevant for emotional state. We report effects in the preregistered subgroups in Appendix D.4. In the body of the paper, we present non-preregistered "exploratory" analysis of subgroups defined by above- vs. below-median values of several moderators that are potentially relevant for emotional state. We present unadjusted p-values for subgroup analyses.

Our first exploratory moderator is the interaction of gender and above- vs. below-median age. (The preregistered moderators for the primary outcomes included median splits of age and gender separately, but not interacted.) This is a key potential moderator because as described in the introduction, some people argue that social media have contributed to the recent decline in young people's mental health, with particular concern about Instagram's effect on young women (Twenge 2017; Engeln et al. 2020; Office of the Surgeon General 2021; Wells et al. 2021; Haidt 2023). The age survey question had coarse response categories; we allocate the median category to the above-median group. This yields 18-34 vs. 35+ and 18-24 vs. 25+ as the age groups in the Facebook and Instagram samples, respectively,.

The second exploratory moderator is baseline use. (This was also a preregistered moderator for the primary outcomes.) This is a key potential moderator because deactivation effectively imposes a larger dose of treatment on heavier users.

The third exploratory moderator is baseline emotional state. This is a key potential moderator because Allcott et al. (2020) find that Facebook deactivation had more beneficial effects for people with worse baseline emotional state, perhaps because they were more vulnerable to unfavorable social comparisons.

The final two exploratory moderators measure political engagement. They are key potential moderators because they speak to mechanisms. If more politically engaged people experience larger effects, this suggests that the effects of deactivation might be smaller outside of an election period or after Meta's decision to reduce political content in news feeds. The fourth potential moderator is baseline political participation, which is the sum of indicators for whether the participant reported doing the following six activities in the past month: (i) attended a protest or rally, (ii) contributed money to a political candidate or organization, (iii) signed an online petition, (iv) tried to convince someone how to vote, (v) posted political messages online, and (vi) talked about politics. (This was also recorded at endline and was preregistered as a primary outcome.) The fifth potential moderator is the count of civic and political pages that the user was following at the beginning of the experiment. This measure only exists for Facebook.

2 Descriptive Statistics

On Facebook, a total of 10.6 million users were invited to the study, 673,388 clicked the invitation, and 43,249 were willing to deactivate, consented to participate, and completed the enrollment survey. Of these, 19,857 completed the baseline survey, could be linked to platform data, and had at least 15 minutes of baseline use per day. This final group is our "primary analysis sample." On Instagram, the analogous numbers are 2.6 million invites, 319,271 clicks,

42,658 enrollment survey completes, and 15,585 participants in the primary analysis sample. See Appendix Table S10 for details.

The fact that less than one percent of the people who were invited to the study completed the experiment underscores that one should be cautious in generalizing results outside our sample. Most of this sample selection is driven by the fact that only a few percent of people click on research study invitations or social media ads. The degree of sample selection in our study is slightly less severe than previous social media deactivation studies, primarily because Meta's research study invitations (which were fixed at the top of users' news feeds) had a higher click-through rate than the standard social media ads used in prior work.[8] In comparison to the full populations of Facebook and Instagram users, our study sample has a higher proportion of users with liberal views and high civic engagement; see Appendix Tables S11 and S12. Our sample weights address sample selection on these and other observables.

Of participants in the Facebook and Instagram primary analysis samples, respectively, 17,802 and 13,480 completed the endline survey. These attrition rates (10 and 13 percent) are relatively low: they are less than the mean of 96 field experiments published in top economics journals that were surveyed in Ghanem, Hirshleifer and Ortiz-Becerra (2022). However, Appendix Table S13 shows that in both experiments, the Deactivation group was about two percentage points more likely to finish endline than the Control group, and this difference is statistically significant in our large sample. Relatedly, Appendix Figure S6 shows that the Deactivation group generally completed endline earlier than the Control group. We assess the possible importance of this differential attrition and response timing below.

The Deactivation and Control groups are balanced on a pre-specified set of demographics at both baseline and endline; see Appendix Tables S14 and S15. In non-preregistered exploratory analysis, we also test for differences between the Deactivation and Control groups in baseline

[8] In Allcott et al. (2020), 1.9 million users were shown ads, of whom 2,897 were randomized. In Asimovic et al. (2021), 365,599 users saw ads, of whom 556 were randomized.

values of the emotional state variables; see Appendix Tables S16 and S17. For Instagram, there are no statistically significant differences. For Facebook, the Deactivation group has statistically significantly worse baseline emotional state. Since there is no evidence of imbalance elsewhere in the experiments, this appears to have occurred by chance. If we did not control for baseline emotional state, this imbalance would bias against our finding that Deactivation improved emotional state. As pre-specified, our regressions control for baseline emotional state, which mitigates that bias. Bruhn and McKenzie (2009) find that after controlling for observables that predict the outcome (which we do), randomizations with chance imbalance are no more likely to generate false hypothesis rejections than those without chance imbalance.

Participants set up passive tracking gradually after the enrollment survey. We limit our passive tracking analyses to participants who have (i) at least two days of baseline data before deactivation began on September 23 and (ii) non-missing data on at least 85 percent of days both during their baseline period and in the post-intervention period between September 23 and November 4. About 29 and 25 percent of the Facebook and Instagram primary analysis samples have valid passive tracking data, and these shares are not statistically different between Deactivation and Control groups.

Appendix B provides descriptive evidence on the time path of emotional state in our sample. Prior work has documented (i) overall negative effects of exposure to politics (Suzuki et al. 2023), (ii) lower average well-being of Democrats relative to Republicans (Simchon et al. 2020), and (iii) short-lived drops in subjective well-being when someone's preferred candidate loses (Pierce et al. 2016). Our data are consistent with all three of these facts. Control group users reported worse emotional state on the endline survey (covering the period leading up to and including election day) than they do at baseline or on a post-endline survey in December. Democrats report worse emotional state than Republicans at baseline, but the gap may have narrowed among people who completed endline after media outlets called the election for

Biden.

3 Empirical Strategy

As we discuss below, not all Deactivation group participants chose to stay deactivated. To estimate the causal effect of deactivation in the presence of this imperfect compliance, we pre-specified that we would use the following instrumental variables regression. We define D_i as a measure of participant i's deactivation compliance, $\boldsymbol{X}_i$ as a vector of controls, ν_s as a vector of randomization stratum indicators, and Y_i as an outcome. The estimating equation is

$$Y_i = \tau D_i + \rho \boldsymbol{X}_i + \nu_s + \varepsilon_i, \tag{1}$$

where we instrument for D_i with a Deactivation group indicator variable T_i. Unless otherwise stated, all analyses below weight observations to make the sample representative of focal platform users with baseline use greater than 15 minutes per day on race, political party, education level, and several measures of platform use. Appendix A.2 further describes the sample weights.

As pre-specified, the control variables $\boldsymbol{X}_i$ are those selected by a lasso regression of the endline outcome Y_i on its baseline value plus a set of demographic variables and baseline political beliefs and behaviors. Deactivation compliance is defined as $D_i = 1 - U_i/\bar{U}_C$, where U_i is the share of days that participant i used the platform (measured by seeing five or more pieces of content) during the September 30 - November 3 treatment period, and $\bar{U}_C$ is the Control group average. $D_i = 1$ for participants who never used the platform, and $D_i = 0$ for those with usage equal to the Control group average. Thus, τ measures the local average treatment effect of never using the platform instead of using the Control group average, for people induced to deactivate by the \$150 payment.

4 Impact Evaluation Results

4.1 First Stage

Before the experiment began, about 90 percent of participants in each experiment used the focal platform on any given day. During the five weeks from September 30-November 3, when only the Deactivation group was being paid to stay deactivated, the Control groups' daily usage rate was again about 90 percent, compared to 15-20 percent in the Deactivation groups. Correspondingly, the first stage coefficients in the Facebook and Instagram samples are 0.871 and 0.893, respectively. See Appendix D.1 for details.

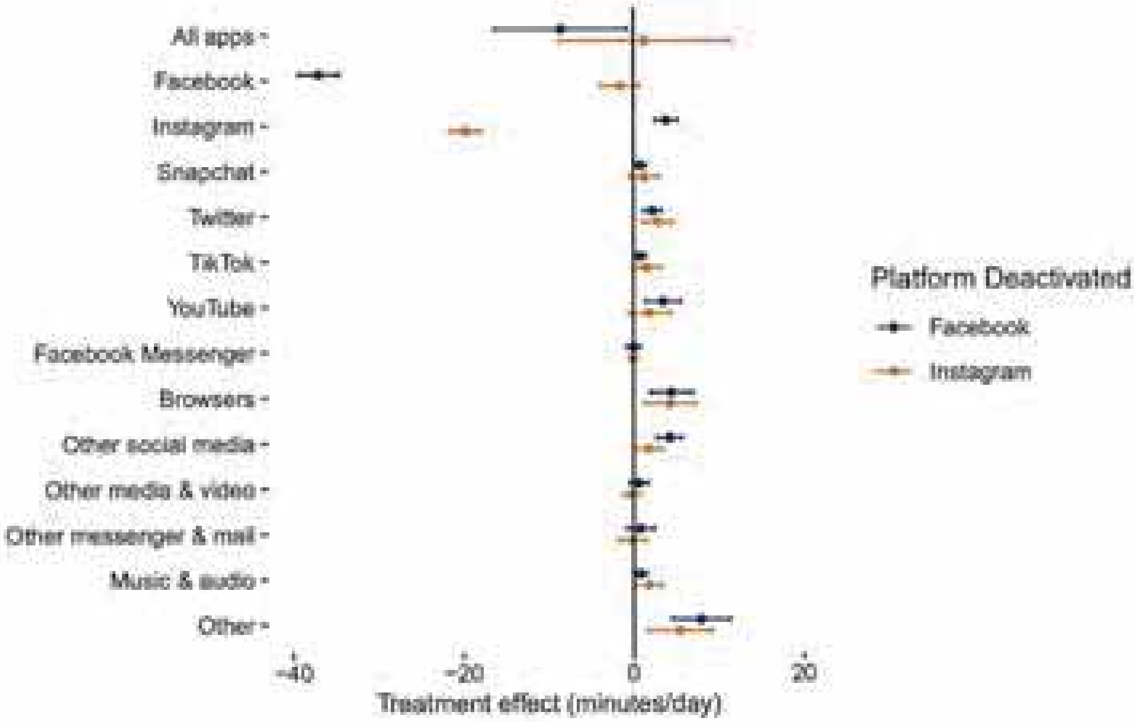
4.2 Substitution to Other Apps

How people reallocate the time gained from deactivation could be a crucial determinant of the effects on emotional state (DellaVigna and La Ferrara 2015; Allcott et al. 2020). For example, if any effects of Facebook or Instagram are from exposure to stressful political content or content that induces unfavorable social comparisons, it would matter if users substitute to other apps with similar content. If any effects are from reduced in-person interactions, it might matter if deactivation reduces overall smartphone screen time.

Figure 1 presents the effects of Facebook and Instagram deactivation on time spent on mobile applications measured in our passive tracking sample. We present the combined effect on all mobile apps (including Facebook and Instagram), then individual effects on seven heavily used social media apps (again including Facebook and Instagram), and finally for all remaining apps grouped into six categories. Since the bottom 13 rows are mutually exclusive and exhaustive, those effects sum to the effect on all apps presented in the first row. For reference, the Facebook and Instagram Control group participants in the passive tracking sample averaged about 51 and 25 minutes per day, respectively, on the Facebook and Instagram mobile apps during the deactivation period; see Appendix Figure S8. These groups averaged 108 and 126

minutes per day, respectively, on all social media apps combined.[9]

Figure 1: Effects of Deactivation on Use of Selected Applications



Note: This figure presents local average treatment effects of Facebook and Instagram deactivation estimated using equation (1). "All apps" is the sum of time spent across all mobile applications. The apps and groups in the next 13 rows are mutually exclusive and exhaustive. The horizontal lines represent 95 percent confidence intervals.

The first row of Figure 1 shows that Instagram deactivation had a small and insignificant effect on total app usage, implying that all of the time participants would have spent on Instagram was substituted to other apps. Facebook deactivation decreased total app usage by an estimated 9 minutes per day. This suggests that while Facebook deactivation increased time spent offline, Instagram deactivation did not. Allcott et al. (2020) report larger substitution to time spent offline based on self-reported data.

[9]Several estimates in this section are slightly different than in Allcott et al. (2024) due to our improved approach to addressing missing passive tracking data.

The second and third rows show that Facebook and Instagram deactivation reduced use of the focal platforms by 37 and 20 minutes per day, respectively. Comparing the coefficients in the first and second rows indicates that around three-quarters of the reduction in Facebook use from deactivation was substituted to other apps.

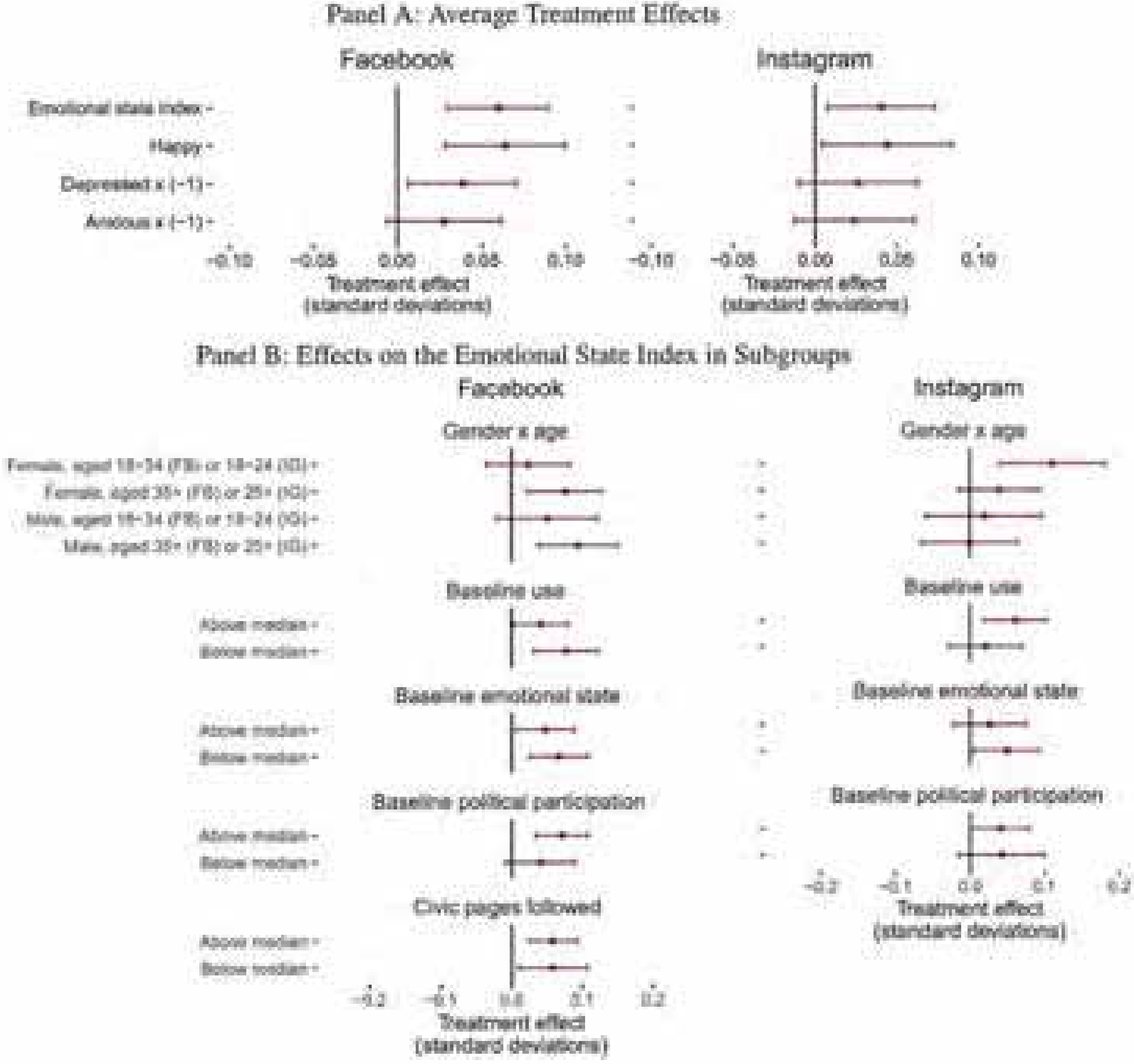
Facebook deactivation increased Instagram use by 4 minutes per day, while Instagram deactivation did not significantly affect Facebook use. The next eleven rows show that Facebook and Instagram deactivation both increased use of Twitter, Snapchat, TikTok, YouTube, web browsers, other social media apps, and other non-categorized apps by a few minutes per day.

Thus, the effects on emotional state that we document below reflect the combined effect of reduced use of the focal platform and increased use of other substitute apps.

4.3 Effects of Deactivation on Emotional State

Average effects. Panel A of Figure 2 reports the local average treatment effects of Facebook and Instagram deactivation on the *emotional state index* and its three components. Appendix D.3 presents all point estimates and p-values. Facebook and Instagram deactivation improved *emotional state index* by 0.060 standard deviations ($p < 0.001$) and 0.041 standard deviations ($p = 0.016$), respectively. The q-values adjusting for multiple hypothesis testing along with the full set of political outcomes considered in Allcott et al. (2024) are 0.002 and 0.139 for Facebook and Instagram; the latter is not statistically significant based on our pre-registered significance threshold of 0.05. Facebook deactivation improved the underlying *happy*, *depressed x (-1)*, and *anxious x (-1)* outcomes by 0.064, 0.039, and 0.028 standard deviations, respectively, with p-values of ¡0.001, 0.018, and 0.110. Instagram deactivation improved those outcomes by 0.044, 0.026, and 0.024 standard deviations, respectively, with p-values of 0.030, 0.156, and 0.205. All four point estimates are smaller for Instagram than for Facebook. In both experiments, the point estimate is largest for *happy* and smallest for *anxious x (-1)*.

Figure 2: Effects of Facebook and Instagram Deactivation on Emotional State



Note: This figure presents local average treatment effects of Facebook and Instagram deactivation estimated using equation 1. The horizontal lines represent 95 percent confidence intervals.

Subgroup analyses. Panel B of Figure 2 presents subgroup analysis for the five emotional state moderators introduced above. Appendix D.4 presents all point estimates and p-values. The first set of results shows differing patterns of age and gender estimates in the two experi-

ments. For Facebook, the point estimates are larger for people over 35 than for younger users. The effects for 35+ vs. 18-34 are statistically different with $p = 0.046$. For Instagram, the estimates are largest for women aged 18-24: an improvement of 0.111 standard deviations ($p =$ 0.002). The point estimates for all other groups are less than half as large and are statistically indistinguishable from zero. The effects on the four age-by-gender subgroups are statistically different with $p = 0.062$. These results are consistent with public concerns described above about the effects of Instagram on young women. Appendix Table S28 shows that the estimates change little when also controlling for baseline use or emotional state, suggesting that the effects on young women are not driven by those correlated factors.

The second and third sets of results show that the effects are not statistically different for above- vs. below-median baseline use or emotional state.

The fourth set of results shows that for both Facebook and Instagram, the point estimates are larger for more politically engaged users, although the two estimates are not statistically different. For Facebook, the effects also do not differ by the count of civic pages followed at baseline. This provides no evidence that the effects of deactivation are related to the political content that circulated during the election period.

Appendix D.4 suggests possible heterogeneity along two other moderators: the point estimates are larger for undecided voters ($p = 0.053$ in a test of equality with non-undecided) and for people without college degrees ($p = 0.058$ in a test of equality with non-college).

4.4 Contextualizing Magnitudes

We benchmark these effect sizes in several ways. First, we compute how far these effects would move people in the distribution of the *emotional state index*. Under the approximation that *emotional state index* is normally distributed, the estimated effects of Facebook or Instagram deactivation would move the median user from the 50th percentile to the 52.4th or 51.6th per-

centile, respectively.

A second benchmark is the effects in their original units, before standardization. Recall that the survey question response options were "All of the time," "Often," "Sometimes," "Rarely," and "Never," which were coded as 4, 3, 2, 1, and 0, respectively. In those original units, Facebook deactivation improved happiness, depression, and anxiety by 0.053, 0.045, and 0.031, respectively. Similarly, Instagram deactivation improved happiness, depression, and anxiety in those original units by 0.037, 0.031, and 0.027, respectively. The average of these six effects is 0.038. This is equivalent to 3.8 percent of people saying they feel happy "often" instead of "sometimes."

A third benchmark is the conditional associations from a regression of baseline *emotional state index* on demographic characteristics, which is reported in Appendix Table S23. The *emotional state index* is 0.09 standard deviations higher for college graduates, 0.16 standard deviations higher for people earning $50,000 to $100,000 than for people earning less than $50,000, 0.23 standard deviations higher for Black and Hispanic people than for other groups, 0.22 standard deviations lower for women than for men, and 0.48 standard deviations higher for Republicans than for Democrats.

As a fourth benchmark, the Control groups report roughly 0.07-0.09 standard deviations worse emotional state at endline than they do at baseline or on a post-endline survey fielded in December. Thus, more than half of the drop in emotional state around the election was eliminated by deactivating Facebook or Instagram.

As a fifth benchmark, the van Agteren et al. (2021) meta-analysis of 419 randomized trials finds that nine types of psychological interventions, such as cognitive behavioral therapy and mindfulness, improve subjective well-being by 0.16 to 0.42 standard deviations. The average across the nine intervention types is 0.27 standard deviations. Thus, the point estimates imply that pre-election Facebook and Instagram deactivation, respectively, improved emotional state

by 15 and 22 percent as much as the average psychological intervention.

As a sixth benchmark, we can compare the effects to the national decline in young people's subjective well-being, which some observers attribute partially or entirely to social media use. To our knowledge, the questions that are closest to our survey questions and have been asked for an extended historical period in the U.S. are the Kessler-6 psychological distress scale asked on the National Survey on Drug Use and Health from 2008-2022.[10] From 2008 to 2022, Kessler-6 scores worsened by 0.37 standard deviations for people aged 18-24. The estimated effect of Instagram deactivation on people aged 18-24 is about 17 percent as large as this time-series change. Of course, this comparison is not informative about the share of the national trend caused by Instagram: the effect of Instagram being adopted in the full U.S. population over a 15-year period could be quite different from the effect of our incremental five-week, pre-election, individual-level deactivation.

Comparison to other experimental estimates. We also compare our estimates to prior estimates of the effects of social media abstention. We limit this comparison to randomized evaluations of social media abstention with abstention periods of at least one week. To find the set of included studies, we carried out a literature search on Google Scholar and PubMed. Appendix Table S30 presents information on the seven other studies that satisfy these inclusion criteria. The table shows that our experiment goes beyond these prior studies in several ways. We have the longest abstention period (5 weeks, compared to a prior maximum of 4 weeks), largest sample size (31,282 compared to a prior maximum of 1,955), and most rigorous enforcement (directly measured by Meta, compared to self-reports in some papers). Only two other studies had a pre-analysis plan. Appendix Table S29 lists an additional 26 randomized experiments that

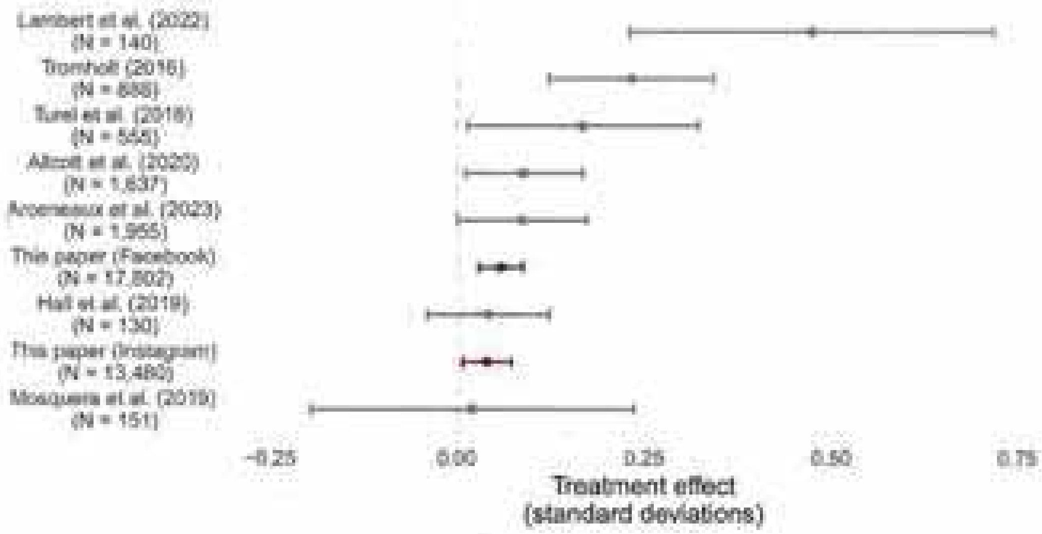
[10]Respondents are asked the share of the time that they felt six negative feelings ("nervous," "hopeless," "restless or fidgety," "so sad or depressed that nothing could cheer you up," "that everything was an effort," and "down on yourself, no good, or worthless"), with answers on the same scale as our questions (a five-point scale from "none of the time" to "all of the time"). We standardize and average these into a single index.

are related but fail our inclusion criteria.

Figure 3 presents the treatment effect estimates from the seven included studies, with effects in units of standard deviations of the outcome variable in the respective study's sample. The smallest effect size is in Mosquera et al. (2020). They find that one week of Facebook abstinence improves self-reported emotional state by a point estimate of 0.02 standard deviations, which is not statistically distinguishable from zero in their sample of 151 people. The largest effect size is in Lambert et al. (2022). They find that one week of social media abstinence improves self-reported emotional state by a statistically significant 0.47 standard deviations, albeit with a very wide confidence interval in their sample of 140 people. Across the seven prior studies, the average confidence interval is 0.28 standard deviations wide, which is 4.60 times larger than the confidence interval around our Facebook estimate. The inverse-variance weighted average effect size for these prior studies is 0.11 standard deviations, which is larger than our estimates.

Braghieri, Levy and Makarin (2022) provide quasi-experimental evidence that Facebook access worsened mental health among college students, leveraging the staggered rollout of the platform to colleges in 2004 and 2005. They estimate a 0.085 standard deviation decline in their mental health index, which is roughly forty percent larger than our point estimate. Their estimate is for a specific subset of our sample population (college students instead of adults 18 and older). Furthermore, the Facebook user experience has changed significantly in the past two decades: for example, there was no news feed, and the user base was over 100 times smaller.

Figure 3: Comparison to Other Experimental Estimates



Note: This figure compares our Facebook and Instagram estimates with other experimental results of social media deactivation by Tromholt (2016), Turel, Cavagnaro and Meshi (2018), Allcott et al. (2020), Mosquera et al. (2020), Hall et al. (2021), Lambert et al. (2022), and Arceneaux et al. (2023). For each paper, we compute treatment effects on the paper's subjective well-being outcomes, in units of standard deviations of the outcome variable in the respective study's sample.

Comparison to non-experimental estimates. We also compare our results to the estimates we would have obtained from non-experimental approaches, which have been used in hundreds of papers (Hancock et al. 2022). In Appendix D.6, we show that both cross-sectional comparisons (controlling for observables) and within-person panel/longitudinal designs give estimates that are biased in unpredictable directions and sometimes have the wrong sign. This highlights the importance of using randomized experiments or credible quasi-experiments for causal inference in this setting.

4.5 Robustness Checks

In Section 2, we documented that in both experiments, the Deactivation group responded earlier and at slightly higher rates than the Control group. Appendix D.7 presents a series of analyses

to diagnose whether this affects the results. We find that excluding control variables, adding controls for endline survey response date, or constructing a sample with balanced endline response rates following Behaghel et al. (2015) all do not substantively affect the results. Lee (2009) bounds exclude zero for *happy* and *emotional state index* in the Facebook experiment and for *happy* in the Instagram experiment, and they rule out negative effects of larger than 0.010 to 0.025 standard deviations for the other five outcomes. Alternative sample weights have limited effects on the results.

5 Conclusion

The relationship between social media use and emotional state is widely debated and of first-order importance for policy. This link is particularly important in the context of an election, where social media may expose users to a range of divisive political content. Existing evidence relies primarily on evidence from time-series and cross-sectional correlations plus a few relatively small randomized experiments. Our experiments are 20 times larger than any previous experiment, the first to consider the effects of Instagram in isolation, and the first to estimate effects in the context of a U.S. presidential election. However, our experiments also have limitations described above, including generalizability, a time-limited intervention, individual-level deactivation, self-reported outcomes, and attrition.

Our estimates suggest that deactivating Facebook or Instagram before the 2020 election improved people's emotional state, although the Instagram effect is not significant at our preregistered threshold after adjusting for multiple hypothesis testing along with the suite of political outcomes in Allcott et al. (2024). The sign of these effects are consistent with public concerns about the effects of social media. However, the estimated effect sizes are smaller than benchmarks such as the effects of psychological interventions, nationwide mental health trends, and previous experimental estimates in smaller samples.

References and Notes

Allcott, Hunt and Matthew Gentzkow, "Social Media and Fake News in the 2016 Election," *Journal of Economic Perspectives*, 2017, *31* (2), 211–236.

_ , Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow, "The Welfare Effects of Social Media," *American Economic Review*, 2020, *110* (3), 629–676.

_ , Matthew Gentzkow, Winter Mason, Arjun Wilkins, Pablo Barberá, Taylor Brown, Juan Carlos Cisneros, Adriana Crespo-Tenorio, Drew Dimmery, Deen Freelon et al., "The Effects of Facebook and Instagram on the 2020 Election: A Deactivation Experiment," *Proceedings of the National Academy of Sciences*, 2024, *121* (21), e2321584121.

American Psychological Association, "Presidential Election a Source of Significant Stress for More Americans Than 2016 Presidential Race," *APA PsycExtra*, 2020.

Arceneaux, Kevin, Martial Foucault, Kalli Giannelos, Jonathan Ladd, and Can Zengin, "The Effects of Facebook Access During the 2022 French Presidential Election: Can We Incentivize Citizens to be Better Informed and Less Polarized?," 2023.

Archie, Ayana, "Montana Becomes the First State to Ban TikTok," Available at `https://www.npr.org/2023/05/18/1176805559/montana-tiktok-ban` 2023.

Asimovic, Nejla, Jonathan Nagler, Richard Bonneau, and Joshua A Tucker, "Testing the Effects of Facebook Usage in an Ethnically Polarized Setting," *Proceedings of the National Academy of Sciences*, 2021, *118* (25), e2022819118.

Behaghel, Luc, Bruno Crépon, Marc Gurgand, and Thomas Le Barbanchon, "Please Call Again: Correcting Nonresponse Bias in Treatment Effect Models," *Review of Economics and Statistics*, 2015, *97* (5), 1070–1080.

Benjamini, Yoav, Abba M Krieger, and Daniel Yekutieli, "Adaptive Linear Step-up Procedures That Control the False Discovery Rate," *Biometrika*, 2006, *93* (3), 491–507.

Braghieri, Luca, Ro'ee Levy, and Alexey Makarin, "Social Media and Mental Health," *American Economic Review*, 2022, *112* (11), 3660–3693.

Brailovskaia, Julia, Fabienne Ströse, Holger Schillack, and Jürgen Margraf, "Less Facebook Use–More Well-being and a Healthier Lifestyle? An Experimental Intervention Study," *Computers in Human Behavior*, 2020, *108*, 106332.

_ , Verena J Swarlik, Georg A Grethe, Holger Schillack, and Jürgen Margraf, "Experimental Longitudinal Evidence for Causal Role of Social Media Use and Physical Activity in COVID-19 Burden and Mental Health," *Journal of Public Health*, 2023, *31* (11), 1885–1898.

Bruhn, Miriam and David McKenzie, "In Pursuit of Balance: Randomization in Practice in Development Field Experiments," *American Economic Journal: Applied Economics*, 2009, *1* (4), 200–232.

Capraro, Valerio, Laura Globig, Steve Rathje, Zach Rausch, A Wormley, Joanne Olson, . . . , Ethan Kross, Jean Twenge, Jonathan Haidt, and Jay J Van Bavel, "A consensus statement on the impact of social media on adolescents' mental health," Unpublished manuscript 2024.

Casey, Katherine, Rachel Glennerster, and Edward Miguel, "Reshaping Institutions: Evidence on Aid Impacts Using a Preanalysis Plan," *The Quarterly Journal of Economics*, 2012, *127* (4), 1755–1812.

Castaño-Pulgarín, Sergio Andrés, Natalia Suárez-Betancur, Luz Magnolia Tilano Vega, and Harvey Mauricio Herrera López, "Internet, Social Media and Online Hate Speech. Systematic Review," *Aggression and Violent Behavior*, 2021, *58*, 101608.

Chetty, Raj, Matthew O Jackson, Theresa Kuchler, Johannes Stroebel, Nathaniel Hendren, Robert B Fluegge, Sara Gong, Federico Gonzalez, Armelle Grondin, Matthew Jacob et al., "Social Capital I: Measurement and Associations with Economic Mobility," *Nature*, 2022, *608* (7921), 108–121.

Chopik, William J, "Associations Among Relational Values, Support, Health, and Well-being Across the Adult Lifespan," *Personal Relationships*, 2017, *24* (2), 408–422.

Cohen, Rachel, Jasmine Fardouly, Toby Newton-John, and Amy Slater, "#BoPo on Instagram: An Experimental Investigation of the Effects of Viewing Body Positive Content on Young Women's Mood and Body Image," *New Media & Society*, 2019, *21* (7), 1546–1564.

Collis, Avinash and Felix Eggers, "Effects of Restricting Social Media Usage on Wellbeing and Performance: A Randomized Control Trial Among Students," *PloS One*, 2022, *17* (8), e0272416.

DellaVigna, Stefano and Eliana La Ferrara, "Economic and Social Impacts of the Media," in "Handbook of Media Economics," Vol. 1, Elsevier, 2015, pp. 723–768.

Dubner, Stephen J, "Is Screen Time as Poisonous as We Think?," Accessed at `https://freakonomics.com/podcast/is-screen-time-as-poisonous-as-we-think`, 2024.

Engeln, Renee, Ryan Loach, Megan N Imundo, and Anne Zola, "Compared to Facebook, Instagram Use Causes More Appearance Comparison and Lower Body Satisfaction in College Women," *Body Image*, 2020, *34*, 38–45.

European Commission, "Commission Opens Formal Proceedings Against Meta Under the Digital Services Act Related to the Protection of Minors on Facebook and Instagram," Avail-

able at https://ec.europa.eu/commission/presscorner/detail/en/ip_24_2664 2024.

Ford, Brett Q, Matthew Feinberg, Bethany Lassetter, Sabrina Thai, and Arasteh Gatchpazian, "The Political is Personal: The Costs of Daily Politics," *Journal of Personality and Social Psychology*, 2023, *125* (1), 1.

Ghanem, Dalia, Sarojini Hirshleifer, and Karen Ortiz-Becerra, "Testing Attrition Bias in Field Experiments," Working Paper September 2022.

González-Bailón, Sandra, David Lazer, Pablo Barberá, Meiqing Zhang, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Deen Freelon, Matthew Gentzkow, Andrew M Guess et al., "Asymmetric Ideological Segregation in Exposure to Political News on Facebook," *Science*, 2023, *381* (6656), 392–398.

Gray, Daniel, Harry Pickard, and Luke Munford, "Election Outcomes and Individual Subjective Wellbeing in Great Britain," *Economica*, 2021, *88* (351), 809–837.

Guess, Andrew M, Neil Malhotra, Jennifer Pan, Pablo Barberá, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Drew Dimmery, Deen Freelon, Matthew Gentzkow et al., "How Do Social Media Feed Algorithms Affect Attitudes and Behavior in an Election Campaign?," *Science*, 2023, *381* (6656), 398–404.

_ , _ , _ , _ , _ , _ , _ , _ , _ , and _ , "Reshares on Social Media Amplify Political News but Do Not Detectably Affect Beliefs or Opinions," *Science*, 2023, *381* (6656), 404–408.

Haidt, Jonathan, "Social Media is a Major Cause of the Mental Illness Epidemic in Teen Girls. Here's the Evidence," *After Babel*, 2023.

Hall, Jeffrey A, Chong Xing, Elaina M Ross, and Rebecca M Johnson, "Experimentally Manipulating Social Media Abstinence: Results of a Four-week Diary Study," *Media Psychology*, 2021, *24* (2), 259–275.

Hancock, Jeff, Sunny Xun Liu, Mufan Luo, and Hannah Mieczkowski, "Psychological Well-being and Social Media Use: A Meta-analysis of Associations Between Social Media Use and Depression, Anxiety, Loneliness, Eudaimonic, Hedonic and Social Well-being," *Anxiety, Loneliness, Eudaimonic, Hedonic and Social Well-Being (March 9, 2022)*, 2022.

Hunt, Melissa G, Rachel Marx, Courtney Lipson, and Jordyn Young, "No More FOMO: Limiting Social Media Decreases Loneliness and Depression," *Journal of Social and Clinical Psychology*, 2018, *37* (10), 751–768.

Huppert, Felicia A, Nic Marks, Andrew Clark, Johannes Siegrist, Alois Stutzer, Joar Vittersø, and Morten Wahrendorf, "Measuring Well-being Across Europe: Description of the ESS Well-being Module and Preliminary Findings," *Social Indicators Research*, 2009, *91*, 301–315.

Imbens, Guido W, "Statistical significance, p-values, and the reporting of uncertainty," *Journal of Economic Perspectives*, 2021, *35* (3), 157–174.

Kimball, Miles S, Collin B Raymond, Jiannan Zhou, Junya Zhou, Fumio Ohtake, and Yoshiro Tsutsui, "Happiness Dynamics, Reference Dependence, and Motivated Beliefs in US Presidential Elections," Technical Report, National Bureau of Economic Research 2024.

Kling, Jeffrey R, Jeffrey B Liebman, and Lawrence F Katz, "Experimental Analysis of Neighborhood Effects," *Econometrica*, 2007, *75* (1), 83–119.

Lambert, Jeffrey, George Barnstable, Eleanor Minter, Jemima Cooper, and Desmond McEwan, "Taking a One-week Break From Social Media Improves Well-being, Depres-

sion, and Anxiety: a Randomized Controlled Trial," *Cyberpsychology, Behavior, and Social Networking*, 2022, *25* (5), 287–293.

Lee, David S. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects," *The Review of Economic Studies*, 2009, *76* (3), 1071–1102.

Meta, "Our Approach to Political Content," https://transparency.meta.com/features/approach-to-political-content 2024.

Mosquera, Roberto, Mofioluwasademi Odunowo, Trent McNamara, Xiongfei Guo, and Ragan Petrie, "The Economic Effects of Facebook," *Experimental Economics*, 2020, *23*, 575–602.

Mummolo, Jonathan and Erik Peterson, "Demand Effects in Survey Experiments: An Empirical Assessment," *American Political Science Review*, 2019, *113* (2), 517–529.

National Conference of State Legislatures, "Social Media and Children Legislation 2023," Available at `https://www.ncsl.org/technology-and-communication/social-media-and-children-2023-legislation` 2024.

New York State Attorney General, "Attorney General James and Multistate Coalition Sue Meta for Harming Youth," Available at `https://ag.ny.gov/press-release/2023/attorney-general-james-and-multistate-coalition-sue-meta-harming-youth` 2023.

Nyhan, Brendan, Jaime Settle, Emily Thorson, Magdalena Wojcieszak, Pablo Barberá, Annie Y Chen, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Drew Dimmery et al., "Like-minded Sources on Facebook are Prevalent but Not Polarizing," *Nature*, 2023, *620* (7972), 137–144.

Odgers, Candice L, "The great rewiring: is social media really behind an epidemic of teenage mental illness?," *Nature*, 2024, *628* (8006), 29–30.

Office of the Surgeon General, *Protecting Youth Mental Health: The U.S. Surgeon General's Advisory*, Washington (DC): US Department of Health and Human Services, 2021.

Ozimek, Phillip and Hans-Werner Bierhoff, "All My Online-friends are Better Than Me–Three Studies About Ability-based Comparative Social Media Use, Self-esteem, and Depressive Tendencies," *Behaviour & Information Technology*, 2020, *39* (10), 1110–1123.

Pierce, Lamar, Todd Rogers, and Jason A Snyder, "Losing Hurts: The Happiness Impact of Partisan Electoral Loss," *Journal of Experimental Political Science*, 2016, *3* (1), 44–59.

Przybylski, Andrew K, Thuy-vy T Nguyen, Wilbert Law, and Netta Weinstein, "Does Taking a Short Break From Social Media Have a Positive Effect on Well-being? Evidence From Three Preregistered Field Experiments," *Journal of Technology in Behavioral Science*, 2021, *6*, 507–514.

Quidt, Jonathan De, Johannes Haushofer, and Christopher Roth, "Measuring and Bounding Experimenter Demand," *American Economic Review*, 2018, *108* (11), 3266–3302.

Reis, Harry T, Andrew W Collins, and Ellen Berscheid, "The Relationship Context of Human Behavior and Development," *Psychological Bulletin*, 2000, *126* (6), 844–872.

Schultz, Alex, "What Do People Actually See on Facebook in the US?," https://about.fb.com/news/2020/11/what-do-people-actually-see-on-facebook-in-the-us/ 2020.

Siegel, Alexandra A, "Online Hate Speech," *Social Media and Democracy: The State of the Field, Prospects for Reform*, 2020, pp. 56–88.

Simchon, Almog, Sharath Chandra Guntuku, Rotem Simhon, Lyle H Ungar, Ran R Hassin, and Michael Gilead, "Political Depression? A Big-data, Multimethod Investigation of Americans' Emotional Response to the Trump Presidency," *Journal of Experimental Psychology: General*, 2020, *149* (11), 2154.

Suzuki, Sara, Lindsay Till Hoyt, Neshat Yazdani, Mariah Kornbluh, Elan C Hope, Melissa J Hagan, Alison K Cohen, and Parissa J Ballard, "Trajectories of Sociopolitical Stress During the 2020 United States Presidential Election Season: Associations with Psychological Well-being, Civic Action, and Social Identities," *Comprehensive Psychoneuroendocrinology*, 2023, *16*, 100218.

Thai, Helen, Christopher G Davis, Wardah Mahboob, Sabrina Perry, Alex Adams, and Gary S Goldfield, "Reducing Social Media Use Improves Appearance and Weight Esteem in Youth With Emotional Distress," *Psychology of Popular Media*, 2023.

Tromholt, Morten, "The Facebook Experiment: Quitting Facebook Leads to Higher Levels of Well-being," *Cyberpsychology, Behavior, and Social Networking*, 2016, *19* (11), 661–666.

Turel, Ofir, Daniel R Cavagnaro, and Dar Meshi, "Short Abstinence From Online Social Networking Sites Reduces Perceived Stress, Especially in Excessive Users," *Psychiatry Research*, 2018, *270*, 947–953.

Twenge, Jean M, *iGen: Why Today's Super-Connected Kids Are Growing Up Less Rebellious, More Tolerant, Less Happy–and Completely Unprepared for Adulthood–and What That Means for the Rest of Us*, Simon and Schuster, August 2017.

Utah Governor's Office, "Utah Sues TikTok Over Child Addiction Harm, Targets "Enmeshment" with its China-Based Parent Company," Available at `https://governor.utah.`

gov/2023/10/10/utah-sues-tiktok-over-child-addiction-harm-targets-enmeshment-with-its-china-based-parent-company/ 2023.

van Agteren, Joep, Matthew Iasiello, Laura Lo, Jonathan Bartholomaeus, Zoe Kopsaftis, Marissa Carey, and Michael Kyrios, "A Systematic Review and Meta-analysis of Psychological Interventions to Improve Mental Wellbeing," *Nature Human Behaviour*, 2021, *5* (5), 631–652.

van Wezel, Marloes MC, Elger L Abrahamse, and Mariek MP Vanden Abeele, "Does a 7-day Restriction on the Use of Social Media Improve Cognitive Functioning and Emotional Well-being? Results From a Randomized Controlled Trial," *Addictive Behaviors Reports*, 2021, *14*, 100365.

Verduyn, Philippe, David Seungjae Lee, Jiyoung Park, Holly Shablack, Ariana Orvell, Joseph Bayer, Oscar Ybarra, John Jonides, and Ethan Kross, "Passive Facebook Usage Undermines Affective Well-being: Experimental and Longitudinal Evidence," *Journal of Experimental Psychology: General*, 2015, *144* (2), 480.

Wasserstein, Ronald L and Nicole A Lazar, "The ASA statement on p-values: context, process, and purpose," *The American Statistician*, 2016, *70* (2), 129–133.

Wells, Georgia, Jeff Horwitz, and Deepa Seetharaman, "Facebook Knows Instagram Is Toxic for Teen Girls, Company Documents Show," https://www.wsj.com/articles/facebook-knows-instagram-is-toxic-for-teen-girls-company-documents-show-11631620739 September 2021. Accessed: 2024-1-13.

Zaller, John and Stanley Feldman, "A simple theory of the survey response: Answering questions versus revealing preferences," *American journal of political science*, 1992, pp. 579–616.

THIS ARTICLE HAS BEEN CORRECTED. SEE LAST PAGE.

AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychology of Popular Media

ISSN: 2689-6567

2025, Vol. 14, No. 1, 1–11
https://doi.org/10.1037/ppm0000536

Limiting Social Media Use Decreases Depression, Anxiety, and Fear of Missing Out in Youth With Emotional Distress: A Randomized Controlled Trial

Christopher G. Davis[1] and Gary S. Goldfield[2]
[1] Department of Psychology, Carleton University
[2] Healthy Active Living and Obesity Research Group, Children's Hospital of Eastern Ontario Research Institute, Ottawa, Canada

Reports demonstrating modest but significant correlations between heavy social media use (SMU) and poorer mental health in youth have led many to suggest that heavy SMU is culpable. Although many youth may not be harmed by heavy SMU, distressed youth may be particularly vulnerable. The aim of this study was to experimentally examine the effects of reducing SMU on smartphones on symptoms of depression, anxiety, fear of missing out (FoMO), and sleep in youth with emotional distress. A randomized controlled trial was used to assign 220 youth aged 17–25 years to either an intervention or control group. The intervention group was asked to reduce smartphone-based SMU to 1 hr/day for 3 weeks while the control group had no SMU restrictions. SMU was objectively measured daily via tracking systems in smartphones. Mental health and sleep were subjectively assessed at baseline and following the 3-week intervention period. Compared to the control group, the intervention group showed significantly greater reductions in symptoms of depression, anxiety, and FoMO, and greater increases in sleep. No effects of gender were detected. Reducing SMU on smartphones to approximately 1 hr/day may be a feasible, inexpensive, and effective method of increasing sleep and reducing symptoms of depression, anxiety, and FoMO among distressed youth.

Public Policy Relevance Statement

A brief 4-week intervention using screen time trackers showed that reducing social media use among heavy social media users reporting symptoms of depression or anxiety yielded significant reductions in symptoms of depression, anxiety, and fear of missing out, and increased hours of sleep relative to a comparable control group. Reducing social media use may be a feasible method of reducing distress among a vulnerable population of heavy social media users.

Keywords: social media, depression, anxiety, fear of missing out, social networking sites

Adolescence and young a… period of development that i… considered to be among the … ment of mental illness due to… and neurobiological change… 20% of youth will be diag… given year (Comeau et al., 2… are the most common fo…

Experimental

…ata (collected during the COVID-19 pandemic) …% and 23% of Canadian youth (18–24 years) …olds for major depressive disorder and generalized …GAD), respectively (Statistics Canada, 2021). …rowing prevalence of anxiety and depressive dis… …th has been the growing use of social media as a …ocial interaction, with approximately 81.3% of …porting moderate-to-heavy (2 hr or more) daily …yinga et al., 2019), and more than 96% using at …edia platform (Sampasa-Kanyinga et al., 2019; …2019). Moreover, data suggest that the vast major… …a use (SMU) in youth occurs on smartphones; …% of American teens report daily access to a … 45% reported being online almost constantly …, 2018). Although SMU is appealing in that it pro… … for social connection and efficient communica… …ss to these platforms combined with the high …tions that effectively tie youth to their smart… …l concerns about the psychological consequences …ong youth.

Several narrative and meta-analytic reviews have indicated small but statistically significant correlations ($r = .11$–$r = .17$) between duration of SMU and depressive symptoms (Orben, 2020). Similarly, a systematic review of 13 studies in youth found that time spent

---

This article was published On…
Christopher G. Davis https…
Gary S. Goldfield https://…
The authors report that there a… that support the findings of thi… Framework at https://osf.io/5c84…

Christopher G. Davis served a… ject administration, and visualiz… supervision and validation. Chris… tributed equally to conceptualiza… and editing, and methodology.

Correspondence concerning this article should be addressed to Gary S. Goldfield, Healthy Active Living and Obesity Research Group, Children's Hospital of Eastern Ontario Research Institute, 401 Smyth Road, Ottawa, ON K1H 8L1, Canada. Email: ggoldfield@cheo.on.ca

THIS ARTICLE HAS BEEN CORRECTED. SEE LAST PAGE.

Psychology of Popular Media

ISSN: 2689-6567

2025, Vol. 14, No. 1, 1–11

Limiting Social Media Use Decreases Depression, Anxiety, and Fear of Missing Out in Youth With Emotional Distress: A Randomized Controlled Trial

Christopher G. Davis[1] and Gary S. Goldfield[2]
[1] Department of Psychology, Carleton University
[2] Healthy Active Living and Obesity Research Group, Children's Hospital of Eastern Ontario Research Institute, Ottawa, Canada

Reports demonstrating modest but significant correlations between heavy social media use (SMU) and poorer mental health in youth have led many to suggest that heavy SMU is culpable. Although many youth may not be harmed by heavy SMU, distressed youth may be particularly vulnerable. The aim of this study was to experimentally examine the effects of reducing SMU on smartphones on symptoms of depression, anxiety, fear of missing out (FoMO), and sleep in youth with emotional distress. A randomized controlled trial was used to assign 220 youth aged 17–25 years to either an intervention or control group. The intervention group was asked to reduce smartphone-based SMU to 1 hr/day for 3 weeks while the control group had no SMU restrictions. SMU was objectively measured daily via tracking systems in smartphones. Mental health and sleep were subjectively assessed at baseline and following the 3-week intervention period. Compared to the control group, the intervention group showed significantly greater reductions in symptoms of depression, anxiety, and FoMO, and greater increases in sleep. No effects of gender were detected. Reducing SMU on smartphones to approximately 1 hr/day may be a feasible, inexpensive, and effective method of increasing sleep and reducing symptoms of depression, anxiety, and FoMO among distressed youth.

Public Policy Relevance Statement

A brief 4-week intervention using screen time trackers showed that reducing social media use among heavy social media users reporting symptoms of depression or anxiety yielded significant reductions in symptoms of depression, anxiety, and fear of missing out, and increased hours of sleep relative to a comparable control group. Reducing social media use may be a feasible method of reducing distress among a vulnerable population of heavy social media users.

Keywords: social media, depression, anxiety, fear of missing out, social networking sites

Adolescence and young adulthood spanning ages 17–25 years is a period of development that is often referred to as youth and is widely considered to be among the most vulnerable periods for the development of mental illness due to the unique social, physical, emotional, and neurobiological changes (Paus et al., 2008). Approximately 20% of youth will be diagnosed with a mental disorder in any given year (Comeau et al., 2019). Depression and anxiety disorders are the most common forms of mental illness with recent epidemiological data (collected during the COVID-19 pandemic) indicating that 36% and 23% of Canadian youth (18–24 years) met clinical thresholds for major depressive disorder and generalized anxiety disorder (GAD), respectively (Statistics Canada, 2021).

Paralleling the growing prevalence of anxiety and depressive disorders among youth has been the growing use of social media as a primary form of social interaction, with approximately 81.5% of Canadian youth reporting moderate-to-heavy (2 hr or more) daily use (Sampasa-Kanyinga et al., 2019), and more than 96% using at least one social media platform (Sampasa-Kanyinga et al., 2019; Statistics Canada, 2019). Moreover, data suggest that the vast majority of social media use (SMU) in youth occurs on smartphones; approximately 95% of American teens report daily access to a smartphone, while 45% reported being online almost constantly (Anderson & Jiang, 2018). Although SMU is appealing in that it provides opportunities for social connection and efficient communication, constant access to these platforms combined with the high volume of notifications that effectively tie youth to their smartphones has sparked concerns about the psychological consequences of heavy SMU among youth.

Several narrative and meta-analytic reviews have indicated small but statistically significant correlations ($r = .11$–$r = .17$) between duration of SMU and depressive symptoms (Orben, 2020). Similarly, a systematic review of 13 studies in youth found that time spent

This article was published Online First April 22, 2024.
Christopher G. Davis https://orcid.org/0000-0003-3851-080X
Gary S. Goldfield https://orcid.org/0000-0001-6216-7824
The authors report that there are no competing interests to declare. The data that support the findings of this study are available in the Open Science Framework at https://osf.io/5z847/.

Christopher G. Davis served as lead for data curation, formal analysis, project administration, and visualization. Gary S. Goldfield served as lead for supervision and validation. Christopher G. Davis and Gary S. Goldfield contributed equally to conceptualization, writing–original draft, writing–review and editing, and methodology.
Correspondence concerning this article should be addressed to Gary S. Goldfield, Healthy Active Living and Obesity Research Group, Children's Hospital of Eastern Ontario Research Institute, 401 Smyth Road, Ottawa, ON K1H 8L1, Canada. Email: ggoldfield@cheo.on.ca

on social media, repeated checking for messages, personal investment on social media, and addictive or problematic SMU was associated with more severe symptoms of depression, anxiety, and psychological distress (Keles et al., 2020).

The relatively weak associations between heavy SMU and psychological distress in youth may indicate that heavy use carries greater risk for some people than others. That is, whereas heavy SMU may not be harmful to well-adjusted, socially integrated youth, similar use may exacerbate distress in vulnerable youth (e.g., youth who feel insecure, have concerns about their body image, are anxious, or are experiencing symptoms of depression or dysphoria). Several mechanisms have been proposed. First, relative to socially integrated youth, vulnerable youth may be more prone to be affected by and targeted with harmful social media content (e.g., teasing, cyberbullying; Sampasa-Kanyinga & Hamilton, 2015; Sampasa-Kanyinga et al., 2018; Sampasa-Kanyinga, Lalande, & Colman, 2020). Second, being exposed to images of celebrities and peers on social media who appear more attractive and seem to live more exciting and interesting lives may lead vulnerable youth to see their own appearance and life as worse by social comparison processes (Pera, 2018; Steers et al., 2014). Additionally, there is emerging evidence to support a third mechanism rooted in displacement theory. This theory was initially developed to explain why television viewing was associated with poorer physical and emotional development (Neuman, 1988), but has been empirically applied to the social media context. Displacement theory posits that engaging in high amounts of time on social media leads to poorer mental health because it displaces time spent in mental health-promoting behaviors (Kraut et al., 1998). Indeed there is empirical support for this theory given epidemiological studies in youth show that heavy (2 hr/day or more) SMU is associated with later bedtime and reduced sleep duration and sleep quality (Sampasa-Kanyinga et al., 2018; Sampasa-Kanyinga, Lalande, & Colman, 2020), which in turn is associated with more severe symptoms of anxiety and depression (Chaput et al., 2016). Also consistent with displacement theory, there is evidence that SMU (and digital media in general) is associated with lower physical activity levels (Sampasa-Kanyinga & Chaput, 2016) and less in-person social interaction, and connectedness with peers and parents (Sampasa-Kanyinga, Goldfield, et al., 2020), factors that are known to protect against the development of psychopathology in youth. It is important to note that these mechanisms are not mutually exclusive.

To date, the evidence for the harmful effects of heavy SMU has not been compelling. As noted by Orben (2020) and Keles et al. (2020), the substantial heterogeneity in methods used across studies combined with the low-quality evidence emanating from predominantly cross-sectional designs limits conclusions drawn from this body of literature, highlighting the need for experimental studies to better understand the psychological effects of SMU.

Few experimental studies on this issue exist. Tromholt (2016) showed that eliminating Facebook use for a period of 1 week significantly improved well-being in Danish adults relative to controls. Turel et al. (2018) likewise demonstrated that short (1 week) periods of abstinence from social media reduced perceived stress, particularly among heavy users. Yet two other studies reported null effects on the well-being of abstaining from SMU for up to 4 weeks (Agadullina et al., 2020; Hall et al., 2021). Thai et al. (2021) found that limiting SMU to 1 hr/day for 3 weeks led to reductions in symptoms of anxiety but not depression in a relatively small sample of youth. In contrast, Hunt et al. (2018) found that reducing SMU on three platforms (Facebook, Snapchat, Instagram) to 10 min/day each for 3 weeks led to a reduction in depressive symptoms among youth. However, they found no intervention effects on anxiety or fear of missing out (FoMO)—the sense of apprehension that one is missing out on pleasant or enjoyable experiences that others are enjoying, a phenomenon shown to motivate greater SMU, with associations with higher distress (Przybylski et al., 2013). Although encouraging, most of these experimental studies have issues that limit their utility. For instance, Agadullina et al. (2020), Hall et al. (2019), and Turel et al. (2018) relied on self-reported compliance to SMU restriction. Self-reported SMU has been shown to be a poor predictor of actual use (Parry et al., 2021). Tromholt (2016), Agadullina et al. (2020), Hall et al. (2019), and Turel et al. (2018) required complete SMU abstinence—a goal that is unrealistic for heavy social media using youth. Hunt et al.'s (2018) restriction to 10 min/day per social media platform may seem overly rigid to youth today, who have more social media options available to them and likely value the freedom to choose how they allocate their SMU when restricted.

The inconsistent findings in the experimental studies and the relatively weak correlations found in cross-sectional studies suggest that not all heavy social media users experience ill effects. Hunt et al. (2018), for instance, found that reducing SMU had greater ameliorative effects on those who were initially depressed relative to those who were not. Of course, this could be attributable to the fact that those with more symptoms have more room to change, but it may also suggest that individuals with elevated distress who are more psychologically vulnerable may be more inclined to make unfavorable social comparisons on social media (Bäzner et al., 2006), to be passive in their use (Verduyn et al., 2017), and/or displace in-person social connections with peers and other mental health-promoting activity in favor of using social media, consistent with displacement theory (Blackwell et al., 2017).

Accordingly, the present study is designed to address the above-noted limitations by experimentally investigating whether reducing objectively measured SMU on smartphones to 1 hr/day for 3 weeks leads to a reduction in depression, anxiety, and FoMO in a large sample of youth with emotional distress. Following Thai et al. (2021), we focus on youth with emotional distress because this is a population that is at greater risk of experiencing the negative effects of heavy use of social media, thus may benefit more from its reduction. In addition to assessing the effect of reducing SMU on indicators of mental health, and guided by displacement theory of how excessive screen time adversely impacts mental health, we also consider the effect that the intervention will have on sleep, given high SMU use is related to later bedtime and reduced sleep duration in youth (Sampasa-Kanyinga et al., 2018; Sampasa-Kanyinga, Lalande, & Colman, 2020), and shorter sleep duration is associated with greater symptoms of anxiety and depression (Chaput et al., 2016).

- Primary hypotheses: Those randomly assigned to voluntarily reduce SMU to 1-hr/day will show a significant reduction in symptoms of depression (Hypothesis 1a [H1a]), anxiety (Hypothesis 1b [H1b]), and FoMO (Hypothesis 1c [H1c]), as well as increases in sleep duration (Hypothesis 1d [H1d]) relative to status quo controls.

Finally, given that women are more likely than men to experience anxiety and depression (Georgiades et al., 2019), combined with some reports of heavier SMU and more adverse associations with mental health among women (Baxter et al., 2014; Keles et al., 2020; Substance Abuse and Mental Health Services Administration, 2017), we consider the extent to which gender (men vs. women) moderates the effect of SMU reduction on mental health outcomes. Many of the correlational studies reviewed above have examined sex differences in the relation between SMU and anxiety/depression, but none of the experimental studies have tested or found gender differences, possibly owing to sample size issues.

- Secondary analyses: Does the effect of reducing social media on depression, anxiety, FoMO, and sleep differ as a function of gender (men vs. women)?

Method

Participants

Undergraduate students enrolled in introductory psychology classes at a Canadian university were recruited to participate in a study entitled "Limiting Social Media Screen-Time on iPhones and Androids." Eligibility requirements were that (a) they were regular social media users (defined as at least 2 hr/day on average), (b) they possessed and regularly used an iPhone (running on iOS 12 or later) or Android smartphone (running on Pie 9 or later), (c) they were between the ages of 17–25 years, and (d) they were experiencing at least two of four symptoms of depression and anxiety as presented on the recruitment notice. All those who were interested were told that each eligible person had a 50% chance of being assigned to the experimental condition where they would be asked to reduce their social media screen time. They were also advised that they would be asked to submit screenshots each morning showing their social media usage for the day before. Although potential participants were aware that the study was about limiting social media screen time (an ethical obligation), they were not aware of our hypotheses until they were debriefed at the study's conclusion.

Participants were recruited over three academic terms (Winter, Summer and Fall of 2021). Two hundred and seventy-nine eligible participants initially enrolled in the study. The Consolidated Standards of Reporting Trials (CONSORT) diagram in Figure 1 summarizes information on dropouts and exclusions. Our analyses are based on a sample of 220 participants (168 women, 50 men, and two indicating "other"). Participants were compensated with grade-raising credit for their introductory psychology class. This study received ethics approval by the authors' institutional review boards and all participants provided informed consent prior to enrollment.

Design

This study used a parallel-group, randomized controlled trial design and was conducted over three semesters spanning January 2021 to December 2021. This study was designed in compliance with CONSORT guidelines for nonpharmacological trials (Boutron et al., 2017). The intervention period lasted for 3 weeks following a 1-week baseline period. Participants had an equal chance of being assigned to either the intervention or the control group using a computer-generated randomization scheme. Participants in the intervention group were instructed to reduce SMU to a maximum of 1 hr/day. This SMU reduction target was based on consultations with our panel of youth with lived experience and guided by the specific, measurable, achievable, relevant, and time-bound (SMART) goal behavioral principles which state that behavior change goals are more likely to be achieved if they are specific, measurable, achievable, relevant, and time-bound. The 1 hr/day SMU goal was also informed by epidemiological data suggesting that using over 1 hr of social media per day is associated with greater emotional distress (Twenge & Campbell, 2019). The social media platforms that were targeted in this study include Facebook, Instagram, TikTok, Snapchat, Twitter, Pinterest, and Tumblr. Messaging apps (e.g., Facebook Messenger, WhatsApp, Reddit, and texts) were not targeted for reduction as research has shown these to be separate domains of social media (Figueroa Jacinto & Arndt, 2018; Kuss & Griffiths, 2017). Participants in the control group monitored their daily SMU and sent daily screenshot but were instructed to use social media as usual throughout the study. Participants in both groups received a daily reminder email for the duration of the study to send their SMU screenshots for the previous day. Furthermore, participants in the intervention group who exceeded the 1 hr/day SMU limit received a reminder email about respecting the SMU limit. Participants were also positively reinforced via email for achieving their SMART goals every time the goal was achieved. These behavioral principles of SMART goal setting, cuing, and immediate positive reinforcement have been empirically shown to be effective components of lifestyle behavior change (Goldfield et al., 2002).

Procedure

As the experiment was conducted during the COVID-19 pandemic, all procedures were conducted virtually. Participants were informed orally (via Zoom) and in writing about the study purpose, its requirements, and potential risks as part of the informed consent process. After confirming eligibility and providing consent, participants were shown how to locate their smartphone's built-in social media tracking summary on their smartphone and were asked to email a test screenshot of it to the secure study inbox. They provided permission for research staff to send daily email reminders to their preferred email address reminding them to submit their daily social media usage screenshot each night. They were then directed to the online baseline questionnaire using a secure data management software, Qualtrics.

For the first 7 days (baseline period), participants were instructed to use social media as usual. On the seventh 7th day, participants were randomly assigned to either the intervention or the control condition. Beginning on the 8th day and lasting until Day 28 (3-week intervention phase), those in the intervention condition were instructed to limit their SMU to 60 min/day, while those assigned to the control group were instructed each day to use social media as usual. On the 28th day, participants were directed to complete postintervention outcome measures online using Qualtrics.

Study materials and data are available for viewing at https://osf.io/5c847/. The study was not preregistered.

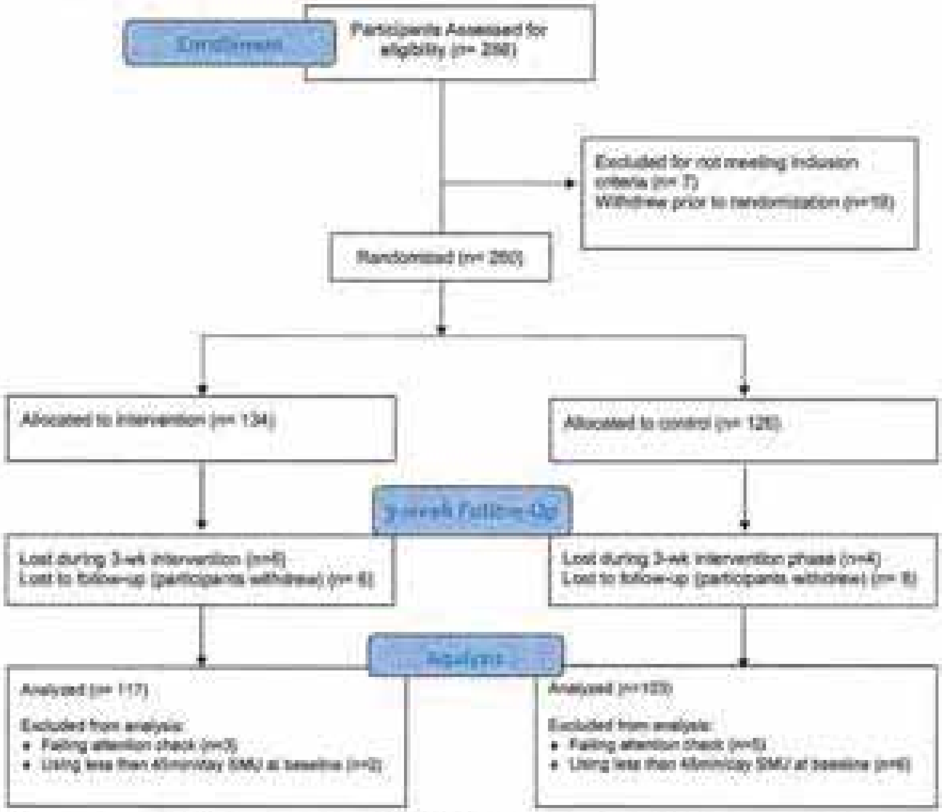
Measures

SMU

Daily SMU via smartphone was objectively assessed using integrated smartphone screen time reports and sent to the study's secured inbox every day for the duration of the study. These applications for

Figure 1

CONSORT Flow Chart of Participants Enrollment, Randomization, Allocations, and Analysis Throughout the Study Timeline



Note. wk = week; SMU = social media use; CONSORT = Consolidated Standards of Reporting Trials. See the online article for the color version of this figure.

iPhones and Android phones enable tracking of time spent on each targeted application, and this objective measurement provides greater reliability compared to self-reported measures of SMU, which are subject to recall bias (Lee et al., 2017; Parry et al., 2021).

Depressive Symptoms

Depressive symptomatology was assessed at the beginning (Day 1) and the end of the experiment (Day 28) with the revised 10-item Center for Epidemiological Studies Depression (CES-D) scale (Andresen et al., 1994; original CES-D by Radloff, 1977). This revised and shortened CES-D has been shown to be a valid and reliable measure of depressive symptoms in a variety of samples (Andresen et al., 1994; González et al., 2017; Irwin et al., 1999) including university students (Bradley et al., 2010). Scores of 10 or more are considered indicative of clinically concerning depression symptoms. Recall that we aimed to target recruitment of youth experiencing emotional distress based on the belief this population would be more vulnerable to the harms of SMU, and may therefore benefit more from its reduction than youth without distress. Although inclusion criteria required all participants to report being distressed, 70% of the sample scored above the threshold of 10 at baseline. Cronbach's α over the two time points averaged .85.

Generalized Anxiety

Generalized anxiety was measured at the beginning and end of the experiment with the GAD-7 (Spitzer et al., 2006). The seven items that make up the GAD-7 are based on *Diagnostic and Statistical Manual of Mental Disorders* (fourth edition, *DSM-IV*) criteria for generalized anxiety. The instrument has been widely used in clinical, general population, and youth and university student samples (e.g., Byrd-Bredbenner et al., 2021; Löwe et al., 2008; Tiirikainen et al., 2019) and has been validated against clinical diagnoses and comparable instruments by Spitzer et al. (2006). A systematic review and meta-analysis of the instrument's diagnostic accuracy against a structured clinical interview indicated that a cut point of 8 had pooled sensitivity and specificity >0.80 (Plummer et al., 2016). In the current study, 58% of participants scored above the clinical threshold of 8 at Time 1. Cronbach's α averaged .89.

FoMO

We used Przybylski et al., (2013) 10-item questionnaire to assess FoMO at the beginning and end of the experiment. This instrument has demonstrated its concurrent validity with positive correlations of scores on the FoMO scale with the level of social media engagement and negative correlations with the level of needs satisfaction, life

satisfaction, and (positive) mood (Przybylski et al., 2013). Moreover, scores on the FoMO scale have also been shown to correlate positively with depression and Internet-communication disorder (Wegmann et al., 2017) and with social anxiety and problematic Facebook use (Dempsey et al., 2019). Items are rated on a 5-point scale where 1 = *not at all true of me* to 5 = *extremely true of me*. We report participants' mean of the 10 items. In the present study, Cronbach's α averaged .86.

Sleep

To assess hours of sleep per night, participants were asked in both the preliminary survey and the follow-up survey the following questions separately for weekdays and weekends: "During the past week, what time have you usually turned out the lights to go to sleep on [weekdays/weekends]?" and "During the past week, what time have you usually woken up in the morning on [weekdays/weekends]?" In the preliminary survey, nine participants did not provide data on their sleep, and in the follow-up survey, 38 did not provide sleep data. Outliers (<2 hr and >14 hr) were replaced so that the minimum hours per night was 2 and maximum was 14. Weekday hours of sleep and weekend hours of sleep were comparable (averaging approximately 8 hr/night; SD = 1.55–1.64) and positively correlated in the baseline survey (r = .51) and in the follow-up survey (r = .68). As such, they were averaged at each time point.

Analysis Strategy

Manipulation Check

A 2 (condition) × 4 (week) mixed analysis of variance (ANOVA) was conducted to compare daily SMU as measured by screenshots of use among participants in each condition during the 4-week study period to evaluate the success of the intervention in limiting SMU.

Primary and Secondary Analysis

To test whether the intervention has effects on primary (depression and anxiety symptoms) and secondary outcomes (FoMO and sleep), 2 × 2 mixed ANOVAs were conducted separately for each outcome. Significant effects were explored using simple effects. We also considered whether effects differed as a function of gender by adding gender (men/women) as a third factor. Effect sizes for each ANOVA are reported in partial eta-squared, with values ranging from 0.01 to 0.05, 0.06 to 0.13, and 0.14 or higher representing small, medium, and large effects, respectively.

Power Analysis

Statistical power was calculated using general linear mixed model power and sample size 3.0.0 (Kreidler et al., 2013) software designed to assess power in longitudinal designs. Power was estimated based on a hypothesized Condition × Time interaction using a repeated measures ANOVA such that those receiving the intervention were expected to decline one third of a standard deviation in our main outcomes (depression and anxiety) whereas controls were expected to maintain their prior level, and anticipating a pre-to-post correlation in dependent variables of r = .6 (stronger test–retest correlations yield greater power). Effect sizes were estimated based on findings of Hunt et al. (2018) and Thai et al. (2021). Assuming a criterion for significance of p = .05, a sample of 200 participants was required to achieve this between group effect size with a power of 0.80.

Results

Participants in both the experimental and control groups reliably provided daily screenshots of the SMU. During the baseline period, 94.5% provided screenshots on all 7 days. In the 21 days following assignment to intervention or control condition, 93.2% provided screenshots on at least 20 days, with rates not differing by condition for either baseline (p = .71) or during intervention (p = .99). Mean daily social media screen time was calculated for each week, with participant's daily mean for that week substituted for any missing data.

Experimental and control groups did not differ significantly at baseline on any of the study variables (see Table 1).

Manipulation Check

A 2 (condition) × 4 (week) ANOVA on average daily social media screen time revealed the expected condition by time (in

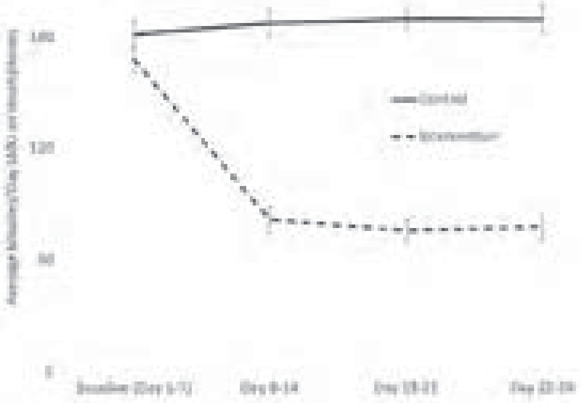
Table 1

Baseline Characteristics of Study Population on Demographic Values and Mental Health Indicators

Variable	Grouped (n = 220)	Intervention (n = 117)	Control (n = 103)	Differences between groups (p)
	Condition			
Gender	50M/168W/2O	25M/92W	25M/76W/2O	.55
Age				.39
17–19 years old (n)	161	84	77	
20–22 years old (n)	37	22	15	
23–25 years old (n)	11	4	7	
Baseline variable	M (SD)	M (SD)	M (SD)	
Depression	14.19 (5.66)	13.96 (5.51)	14.45 (5.84)	.52
Anxiety	9.86 (6.18)	9.43 (6.42)	10.34 (5.86)	.28
FoMO	2.68 (.84)	2.65 (.86)	2.72 (.82)	.59
Sleep (hour/day)	8.35 (1.39)	8.27 (1.49)	8.43 (1.27)	.39

Note. Missing age data of 11 participants due to technical issues. Missing sleep data for nine participants due to nonresponse. M = men; W = women; O = other; FoMO = fear of missing out.

Figure 2
Total Daily SMU Over Time by Condition



Note. Error bars represent standard errors. SMU = social media use.

weeks) interaction, $F(3, 648) = 94.048$, $p < .001$, $\eta_p^2 = .255$. Simple effects indicated no difference between intervention and control groups during the baseline period (days 1–7; $p = .197$), but significant differences by condition in each of the subsequent weeks (all $p < .001$), with those in the intervention condition averaging 78.25 min/day (reducing their daily SMU by approximately 50%) whereas those in the control condition averaging 188.76 min/day (see Figure 2).

Main Analyses

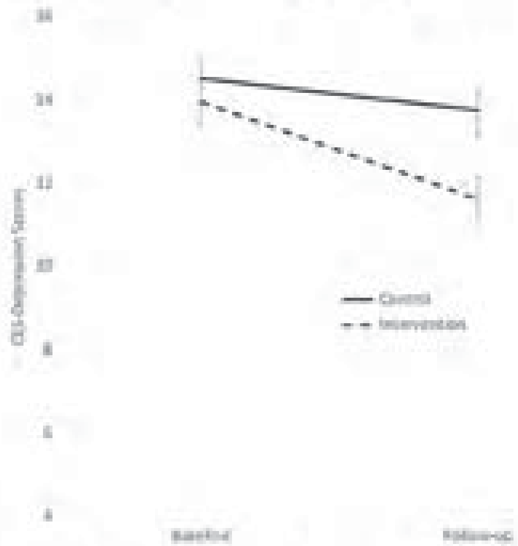
Effects of Reducing Social Media on Depressive Symptoms (H1a)

The ANOVA on depression indicated a marginal main effect of condition, $F(1, 217) = 3.63$, $p = .058$, $\eta_p^2 = .016$; a significant main effect of time, $F(1, 217) = 21.42$, $p < .001$, $\eta_p^2 = .090$; and a significant condition by time interaction, $F(1, 217) = 5.35$, $p = .022$, $\eta_p^2 = .024$. Simple effects indicated that whereas the control group did not decline significantly (decrease of 0.79, $p = .115$), the intervention group declined significantly in levels of depression (decrease of 2.36 points, $p < .001$; see Figure 3). Adding gender as a factor to the model yielded no significant effects involving gender (men vs. women) and did not significantly alter any of the effects described above.

Effects of Reducing Social Media on Anxiety Symptoms (H1b)

The ANOVA on anxiety indicated a significant main effect of condition, $F(1, 216) = 4.33$, $p = .039$, $\eta_p^2 = .020$; of time, $F(1, 216) = 37.17$, $p < .001$, $\eta_p^2 = .147$; and an interaction of condition by time, $F(1, 216) = 5.99$, $p = .015$, $\eta_p^2 = .027$. Simple effects indicated that whereas both groups declined over time, the decrease was greater for those assigned to the intervention (decrease of 2.35 points, $p < .001$) relative to controls (decrease of 1.01 points, $p = .013$; see Figure 4). Adding gender as a factor to the model yielded no significant effects involving gender and did not significantly alter any of the effects described above.

Figure 3
Effect of Reducing SMU on Symptoms of Depression

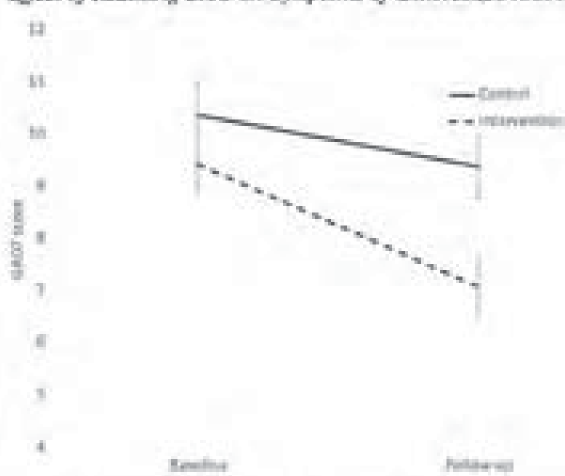


Note. Error bars represent standard errors. SMU = social media use; CES = Center for Epidemiological Studies.

Effects of Reducing Social Media on FoMO (H1c)

The ANOVA on FoMO indicated no significant main effect of condition, $F(1, 217) = 2.34$, $p = .128$, $\eta_p^2 = .011$, a significant effect of time, $F(1, 217) = 44.60$, $p < .001$, $\eta_p^2 = .170$, and a significant interaction of condition by time, $F(1, 217) = 3.95$, $p = .048$, $\eta_p^2 = .018$. Simple effects indicated that whereas both groups declined over time, the decrease was greater for those assigned to the intervention ($p < .001$) relative to controls ($p = .002$; see Figure 5). Adding gender as a factor to the model yielded no

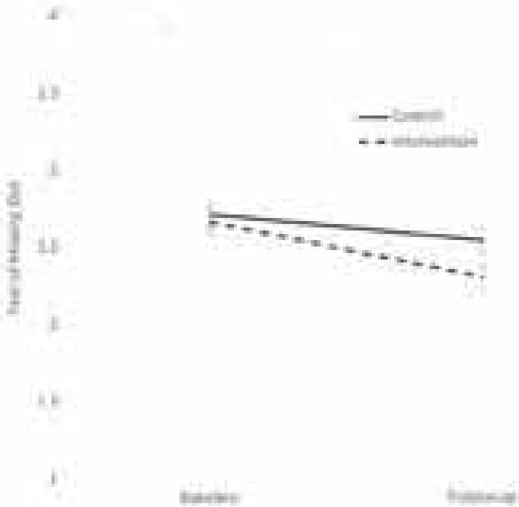
Figure 4
Effect of Reducing SMU on Symptoms of Generalized Anxiety



Note. Error bars represent standard errors. SMU = social media use; GAD-7 = generalized anxiety disorder 7.

Figure 5

Effect of Reducing SMU on Fear of Missing Out



Note. Error bars represent standard errors. SMU = social media use.

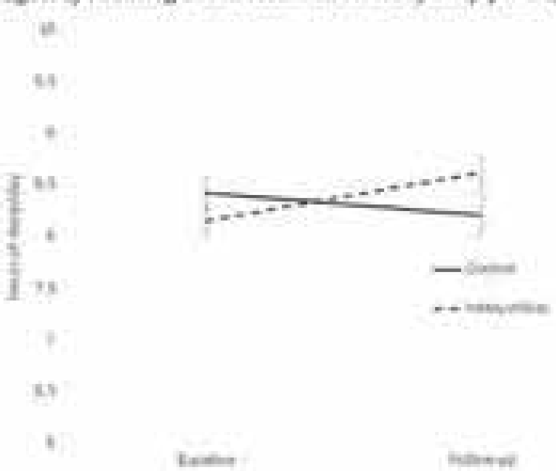
significant effects involving gender and did not significantly alter any of the effects described above.

Effects of Reducing SMU on Sleep (H1d)

The ANOVA on sleep yielded a nonsignificant main effect of time, $F(1, 176) = 1.15$, $p = .284$, $\eta_p^2 = .007$; a nonsignificant main effect of condition, $F(1, 176) = 0.14$, $p = .709$, $\eta_p^2 = .001$; and a significant interaction, $F(1, 176) = 9.52$, $p = .002$, $\eta_p^2 = .051$. Simple effects analyses indicated that whereas hours of sleep declined for control participants by approximately 15 min per night ($p = .179$), sleep increased for those in the intervention condition by about 30 min per night ($p = .002$; see Figure 6). Adding gender as a factor to the model yielded no significant effects involving gender and did not significantly alter any of the effects described above.

Figure 6

Effect of Reducing Social Media on Hours of Sleep per Day



Note. Error bars represent standard errors.

Discussion

As hypothesized, experimentally reducing SMU in youth with emotional distress to approximately 1 h/day for 3 weeks led to significant reductions in anxiety and depression symptoms relative to self-monitoring controls who had unrestricted access to SMU. These findings add experimental evidence showing a clear causal link between SMU and mental health, consistent with a growing body of predominantly cross-sectional evidence showing that heavy use of social media may be psychologically harmful to youth (Twenge & Campbell, 2019; Woods & Scott, 2016). Unlike previous studies, however, we limited our participant pool to those who were currently reporting symptoms of anxiety or depression on the assumption that reducing use is most likely to be salubrious among those with preexisting symptoms. Youth who are not experiencing distress may not be as affected by heavy SMU, and—as Hunt et al. (2018) observed—reducing SMU among this group may render smaller mental health benefits. Moreover, this study found experimental evidence to support the displacement hypothesis in explaining how reducing social media may have a favorable impact on mental health. Specifically, reducing SMU led to increased sleep, and greater sleep is associated with lower anxiety and depression in youth (Chaput et al., 2016).

To our knowledge, our study is the first to show that SMU reduction to approximately 1-h/day for 3 weeks led to a decrease in both anxiety and depressive symptoms in youth with emotional distress, whereas previous experimental studies in healthy populations have shown somewhat inconsistent findings (Apaolaza et al., 2020; Hall et al., 2019; Hunt et al., 2018; Tromholt, 2016). There are many potential explanations for these findings. Youth with emotional distress tend to be less engaged with peers and organized activities (Steger & Kashdan, 2009), report more social isolation (Achterbergh et al., 2020), unfavorable social comparisons (Bäzner et al., 2006) and may use SMU to stay connected with peers to compensate for a lack of offline interpersonal relationships (Blackwell et al., 2017). This may make them more vulnerable to the psychologically harmful elements of heavy SMU exposure that result from the preponderance of portrayals of online profiles and posts that over-represent positive experiences, photo-edited pictures, and displays of high number of friends/followers and "likes" as a reflection of perceived online popularity and elicit envy (Pera, 2018). Thus, reducing time spent using social media may confer psychological benefits by reducing exposure to these "toxic" elements and unfavorable psychological comparison processes, which have been noted to be more frequent and harmful in SMU contexts as compared to offline environments (Lin et al., 2016; Walther et al., 2011). Regardless of the mechanisms, our findings show widespread psychological benefits of reducing SMU in a vulnerable population of youth with emotional distress.

To the extent that people use social media to stay connected with peers (especially during a pandemic when this study was conducted), one might expect that limiting one's time on social media might leave one feeling left out and isolated. Some have argued that such FoMO keeps people tied to social media (Przybylski et al., 2013), but also that high SMU leads to greater FoMO (Oberst et al., 2017). This highlights the bidirectional nature of these relationships. The drive to use smartphones is very powerful, evidenced by studies showing social media possesses greater reinforcing properties than palatable snack foods (O'Donnell & Epstein, 2019) and evokes the release of

dopamine, with concomitant activation in brain regions that are implicated in the development of drug and alcohol addiction (Montag et al., 2017; Sariyska et al., 2018). Similarly, it has been demonstrated that the intermittent and unlimited reinforcement from receiving SMU notifications leads to habitual SMU as a means to prevent FoMO (Griffiths, 2018). Given these reinforcement processes inherent in SMU combined with the fact that we did not instruct participants to turn off their SMU notifications, it would not be unreasonable to predict that limiting access to SMU would increase FoMO, but ironically our study showed the opposite; that reducing SMU by about 90 min/day (approximate 50% reduction from baseline) for 3 weeks led to a greater reduction in FoMO relative to controls. This is a novel finding that has important public health implications given the high prevalence of excessive SMU in youth, and the reliable relationship between FoMO and emotional distress (Przybylski et al., 2013). Interestingly, our results are consistent with studies using an abstinence model that show that whereas short-term (24-hr) abstinence from SMU may temporarily increase perceived FoMO and emotional distress (Roberts & Koliska, 2014), 7-day abstinence tends to reduce FoMO and increase social connection (Brown & Kuss, 2020). This pattern of results is also consistent with addiction research indicating that the heightened distress initially experienced upon cessation or marked reduction in use (i.e., withdrawal) decreases over time with adaptation when individuals learn to meet their psychological needs in other ways (Brown & Kuss, 2020; Turel, 2015). Reducing the exposure of SMU to a moderate amount of use (about 1-hr/day) may provide an optimal balance between allowing enough SMU to feel connected to peers and meet psychological needs while sufficiently reducing exposure to the harmful effects of SMU (e.g., unfavorable social comparisons), and/or perhaps offering time for healthier pursuits, leading to reduced FoMO and emotional distress.

According to the displacement hypothesis (Neuman, 1988), spending large amounts of time on social media (or screens in general) displaces time spent on mental health-promoting behaviors like sleep, physical activity, time in nature, and recreational activities and hobbies (Guerrero et al., 2019; Nie, 2001). Until now, the evidence for this has relied on data from cross-sectional studies showing that SMU is associated with delayed bedtime, shorter sleep duration, and poorer sleep quality (Levenson et al., 2017; Sampasa-Kanyinga et al., 2018; Sampasa-Kanyinga, Lalande, & Colman, 2020; Scott et al., 2019; Woods & Scott, 2016). Additionally, systematic reviews indicate that shorter sleep duration is associated with greater anxiety and depression in youth (Chaput et al., 2016) and adults (Ross et al., 2020). To our knowledge, our study is the first to experimentally demonstrate that reducing SMU by about 90 min/day led to significant increases—30 min/night—in sleep, whereas the control group showed a reduction of 15 min/night over the course of the study, for a relative group difference of 45 min/night. This suggests that one possible mechanism by which reducing SMU may lead to reductions in emotional distress is through increased sleep, consistent with displacement theory.

Gender Differences

We examined how gender might moderate intervention effects given studies showing that women are at greater risk of anxiety and depression, tend to spend more time in SMU, and may be at greater risk of psychological harm from heavier SMU compared to men (Keles et al., 2020). We found no gender differences in the effect of limiting social media on the psychological outcomes of interest, possibly due limited power resulting from the fact that only 23% of the sample was comprised of men. This gender breakdown was somewhat expected given there are more women than men in undergraduate psychology, and women are more likely to experience emotional distress, and are heavier social media users than are men. That we found no gender differences suggests that whereas young women may be at greater risk, young men and women derive comparable psychological benefits from SMU reduction.

Strengths and Limitations

This study has many methodological strengths and weaknesses that warrant mention. Study strengths include the randomized controlled trial design with an active control group that controlled for self-monitoring of SMU, an objective measure of SMU on smartphones, good compliance to the intervention, and validated measures of mental health, all of which strengthen the internal validity of the findings. In addition, we implemented the intervention virtually at very little cost, enhancing the public health implications of the findings. Moreover, we targeted a population of youth presenting with emotional distress during a critical developmental period that puts them at risk for lifelong mental illness, making results more clinically impactful.

These strengths are balanced by several weaknesses. The intervention period only lasted 3 weeks and thus represents a proof of principle type study. Future studies should investigate whether reducing SMU over longer periods of time is a feasible and effective way for producing sustained mental health benefits in youth. Although most participants were quite compliant with the SMU restriction on their smartphones, we could not objectively clamp SMU, thus some participants in the intervention group greatly exceeded the 1-hr daily goal. Nevertheless, the intervention group still reduced SMU by approximately 50%, on average, from baseline. It should be recognized that the 1 hr/day goal was derived by cross-sectional research showing more than 1 hr of digital media use is associated with greater distress in a dose-response manner (Twenge & Campbell, 2019), this target is somewhat arbitrary, but it served as a specific, viable goal that participants could measure themselves against as opposed to something vague like "reduce to 50% of your typical SMU." Although participants did not always meet the target, they did cut their SMU substantially and consequently were likely more mindful of how they used their time on social media. Finally, our sample comprised university students who were willing to participate in a study where they knew there was a 50% chance that they would be asked to limit their SMU; they had some interest in reducing their SMU. An unselected sample may be less inclined to comply with requests to limit their use, and thus may not realize the mental health benefits.

Future Research Directions

Future research should assess the mechanisms through which reductions in SMU improve mental health. Several putative mechanisms have been proposed, including reductions in unfavorable social comparisons, reduced exposure to harmful content, and more frequent and enhanced in-person social interactions, with evidence from the current study suggestive of increased sleep. As such, results are consistent with displacement theory, but a more rigorous

test of this theory is warranted by incorporating many other health-promoting behaviors beyond sleep that could be displaced by high SMU, and by extension, substituted for when SMU is constrained.

Whereas our study shows that cutting down on SMU reduces FoMO, anxiety, and depressive symptoms, and increases sleep time, we also observed reductions in FoMO and anxiety in the control group, albeit to a lesser extent. Although we do not have evidence to account for this, it may be that these reductions are due to the control group monitoring their daily use (i.e., being made aware daily through submitting screenshots of their SMU how much time they are spending on such platforms). Another possibility has been suggested by Shrout et al. (2018), who have documented the tendency for participants in longitudinal studies to slightly elevate their initial reports, particularly on measures assessing affect. They make the case that it is more likely an initial elevation rather than attenuation at follow-up. Regardless, these trends in the control group highlight the importance of using randomized controlled designs in intervention research.

Finally, our study only targeted reductions in the duration of SMU, but research shows that how people use social media (active vs. passive, open chats vs. closed, number and type of platforms used, etc.) shows differential associations with mental health (Thorisdottir et al., 2019; Tromholt, 2016; Verduyn et al., 2017), so both quantity and quality of SMU should be taken into account when designing intervention studies.

Conclusion

To our knowledge, our results are the first to show that among youth with emotional distress, reducing SMU by about 50% from baseline produced significant reductions in symptoms of depression, anxiety, and FoMO. In addition, we also found that SMU reduction led to a significant increase in sleep, consistent with displacement theory. These beneficial effects of SMU were not moderated by gender, suggesting that men and women with emotional distress derive comparable psychological benefits from reducing SMU. These findings suggest that reducing SMU may represent a feasible, affordable, and effective strategy that should be considered for inclusion in the comprehensive management of anxiety and depression in youth with emotional distress, a high-risk population for chronic mental illness. Future research using well-controlled experimental designs is needed to empirically determine whether displacement theory provides a more comprehensive mechanistic understanding relative to other competing theories (e.g., social comparison theory) on how reducing SMU confers mental health benefits.

References

Achterbergh, L., Pitman, A., Birken, M., Pearce, E., Sno, H., & Johnson, S. (2020). The experience of loneliness among young people with depression: A qualitative meta-synthesis of the literature. *BMC Psychiatry*, *20*(1), Article 415. https://doi.org/10.1186/s12888-020-02818-3

Agadullina, E. R., Lovakov, A., & Koshelnikova, N. V. (2020). Does quitting social networks change feelings of loneliness among freshmen? An experimental study. *Journal of Applied Research in Higher Education*, *13*(1), 149–163. https://doi.org/10.1108/JARHE-11-2019-0283

Anderson, M., & Jiang, J. (2018, May). *Monica Anderson and Jingjing Jiang*. Pew Research. https://assets.pewresearch.org/wp-content/uploads/sites/14/2018/05/31102617/PI_2018.05.31_TeensTech_FINAL.pdf

Andresen, E. M., Malmgren, J. A., Carter, W. B., & Patrick, D. L. (1994). Screening for depression in well older adults: Evaluation of a short form of the CES-D (Center for Epidemiologic Studies Depression Scale). *American Journal of Preventive Medicine*, *10*(2), 77–84. https://doi.org/10.1016/S0749-3797(18)30622-6

Baxter, A. J., Scott, K. M., Ferrari, A. J., Norman, R. E., Vos, T., & Whiteford, H. A. (2014). Challenging the myth of an "epidemic" of common mental disorders: Trends in the global prevalence of anxiety and depression between 1990 and 2010. *Depression and Anxiety*, *31*(6), 506–516. https://doi.org/10.1002/da.22230

Bäzner, E., Brömer, P., Hammelstein, P., & Meyer, T. D. (2006). Current and former depression and their relationship to the effects of social comparison processes. Results of an internet based study. *Journal of Affective Disorders*, *93*(1–3), 97–103. https://doi.org/10.1016/j.jad.2006.02.017

Blackwell, D., Leaman, C., Tramposch, R., Osborne, C., & Liss, M. (2017). Extraversion, neuroticism, attachment style and fear of missing out as predictors of social media use and addiction. *Personality and Individual Differences*, *116*, 69–72. https://doi.org/10.1016/j.paid.2017.04.039

Boutron, I., Altman, D. G., Moher, D., Schulz, K. F., Ravaud, P., & CONSORT NPT Group. (2017). CONSORT statement for randomized trials of nonpharmacologic treatments: A 2017 update and a CONSORT extension for nonpharmacologic trial abstracts. *Annals of Internal Medicine*, *167*(1), 40–47. https://doi.org/10.7326/M17-0046

Bradley, K. L., Bagnell, A. L., & Brannen, C. L. (2010). Factorial validity of the Center for Epidemiological Studies Depression 10 in adolescents. *Issues in Mental Health Nursing*, *31*(6), 408–412. https://doi.org/10.3109/01612840903484105

Brown, L., & Kuss, D. J. (2020). Fear of missing out, mental wellbeing, and social connectedness: A seven-day social media abstinence trial. *International Journal of Environmental Research and Public Health*, *17*(12), Article 4566. https://doi.org/10.3390/ijerph17124566

Byrd-Bredbenner, C., Eck, K., & Quick, V. (2021). GAD-7, GAD-2, and GAD-mini: Psychometric properties and norms of university students in the United States. *General Hospital Psychiatry*, *69*, 61–66. https://doi.org/10.1016/j.genhosppsych.2021.01.002

Chaput, J.-P., Gray, C. E., Poitras, V. J., Carson, V., Gruber, R., Olds, T., Weiss, S. K., Connor Gorber, S., Kho, M. E., Sampson, M., Belanger, K., Eryuzlu, S., Callender, L., & Tremblay, M. S. (2016). Systematic review of the relationships between sleep duration and health indicators in school-aged children and youth. *Applied Physiology, Nutrition, and Metabolism*, *41*(6 Suppl. 3), S266–S282. https://doi.org/10.1139/apnm-2015-0627

Comeau, J., Georgiades, K., Duncan, L., Wang, L., & Boyle, M. H. (2019). Changes in the prevalence of child and youth mental disorders and perceived need for professional help between 1983 and 2014: Evidence from the Ontario Child Health Study. *The Canadian Journal of Psychiatry*, *64*(4), 256–264. https://doi.org/10.1177/0706743719830035

Dempsey, A. E., O'Brien, K. D., Tiamiyu, M. F., & Elhai, J. D. (2019). Fear of missing out (FoMO) and rumination mediate relations between social anxiety and problematic Facebook use. *Addictive Behaviors Reports*, *9*, Article 100150. https://doi.org/10.1016/j.abrep.2018.100150

Figueroa Jacinto, R., & Arndt, S. (2018). Prevalence of the use of mobile instant messaging applications: The need to assess their usage in human factors investigations. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *62*(1), 406–410. https://doi.org/10.1177/1541931218621093

Georgiades, K., Duncan, L., Wang, L., Comeau, J., & Boyle, M. H. (2019). Six-month prevalence of mental disorders and service contacts among children and youth in Ontario: Evidence from the 2014 Ontario Child Health Study. *Canadian Journal of Psychiatry*, *64*(4), 246–255. https://doi.org/10.1177/0706743719830024

Goldfield, G. S., Raynor, H. A., & Epstein, L. H. (2002). Treatment of pediatric obesity. In T. A. Wadden & A. J. Stunkard (Eds.), *Handbook of obesity treatment*. (pp. 532–555). Guilford Press.

González, P., Nuñez, A., Merz, E., Brintz, C., Weitzman, O., Navas, E. L., Camacho, A., Buelna, C., Penedo, F. J., Wassertheil-Smoller, S., Perreira, K., Isasi, C. R., Choca, J., Talavera, G. A., & Gallo, L. C. (2017). Measurement properties of the Center for Epidemiologic Studies Depression Scale (CES-D 10): Findings from HCHS/SOL. *Psychological Assessment, 29*(4), 372–381. https://doi.org/10.1037/pas0000330

Griffiths, M. D. (2018). Adolescent social networking: How do social media operators facilitate habitual use? *Education and Health, 36*(3), 66–69.

Guerrero, M. D., Barnes, J. D., Chaput, J.-P., & Tremblay, M. S. (2019). Screen time and problem behaviors in children: Exploring the mediating role of sleep duration. *International Journal of Behavioral Nutrition and Physical Activity, 16*(1), Article 105. https://doi.org/10.1186/s12966-019-0862-x

Hall, J. A., Johnson, R. M., & Ross, E. M. (2019). Where does the time go? An experimental test of what social media displaces and displaced activities' associations with affective well-being and quality of day. *New Media & Society, 21*(3), 674–692. https://doi.org/10.1177/1461444818804775

Hall, J. A., Xing, C., Ross, E. M., & Johnson, R. M. (2021). Experimentally manipulating social media abstinence: Results of a four-week diary study. *Media Psychology, 24*(2), 259–275. https://doi.org/10.1080/15213269.2019.1688171

Hunt, M. G., Marx, R., Lipson, C., & Young, J. (2018). No more FOMO: Limiting social media decreases loneliness and depression. *Journal of Social and Clinical Psychology, 37*(10), 751–768. https://doi.org/10.1521/jscp.2018.37.10.751

Irwin, M., Artin, K. H., & Oxman, M. N. (1999). Screening for depression in the older adult: Criterion validity of the 10-item Center for Epidemiological Studies Depression Scale (CES-D). *Archives of Internal Medicine, 159*(15), 1701–1704. https://doi.org/10.1001/archinte.159.15.1701

Keles, B., McCrae, N., & Grealish, A. (2020). A systematic review: The influence of social media on depression, anxiety and psychological distress in adolescents. *International Journal of Adolescence and Youth, 25*(1), 79–93. https://doi.org/10.1080/02673843.2019.1590851

Kraut, R., Patterson, M., Lundmark, V., Kiesler, S., Mukophadhyay, T., & Scherlis, W. (1998). Internet paradox: A social technology that reduces social involvement and psychological well-being? *American Psychologist, 53*(9), 1017–1031. https://doi.org/10.1037/0003-066X.53.9.1017

Kreidler, S. M., Muller, K. E., Grunwald, G. K., Ringham, B. M., Coker-Dukowitz, Z. T., Sakhadeo, U. R., Barón, A. E., & Glueck, D. H. (2013). GLIMMPSE: Online power computation for linear models with and without a baseline covariate. *Journal of Statistical Software, 54*(10), Article 110. https://doi.org/10.18637/jss.v054.i10

Kuss, D., & Griffiths, M. (2017). Social networking sites and addiction: Ten lessons learned. *International Journal of Environmental Research and Public Health, 14*(3), Article 311. https://doi.org/10.3390/ijerph14030311

Lee, H., Ahn, H., Nguyen, T. G., Choi, S.-W., & Kim, D. J. (2017). Comparing the self-report and measured smartphone usage of college students: A pilot study. *Psychiatry Investigation, 14*(2), Article 198. https://doi.org/10.4306/pi.2017.14.2.198

Levenson, J. C., Shensa, A., Sidani, J. E., Colditz, J. B., & Primack, B. A. (2017). Social media use before bed and sleep disturbance among young adults in the United States: A nationally representative study. *Sleep, 40*(9), Article zsx113. https://doi.org/10.1093/sleep/zsx113

Lin, L. Y., Sidani, J. E., Shensa, A., Radovic, A., Miller, E., Colditz, J. B., Hoffman, B. L., Giles, L. M., & Primack, B. A. (2016). Association between social media use and depression among US young adults. *Depression and Anxiety, 33*(4), 323–331. https://doi.org/10.1002/da.22466

Löwe, B., Decker, O., Müller, S., Brähler, E., Schellberg, D., Herzog, W., & Herzberg, P. Y. (2008). Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. *Medical Care, 46*(3), 266–274. https://doi.org/10.1097/MLR.0b013e318160d093

Montag, C., Markowetz, A., Blaszkiewicz, K., Andone, I., Lachmann, B., Sariyska, R., Trendafilov, B., Eibes, M., Kolb, J., Reuter, M., Weber, B., & Markett, S. (2017). Facebook usage on smartphones and gray matter volume of the nucleus accumbens. *Behavioural Brain Research, 329*, 221–228. https://doi.org/10.1016/j.bbr.2017.04.035

Neuman, S. B. (1988). The displacement effect: Assessing the relation between television viewing and reading performance. *Reading Research Quarterly, 23*(4), 414–440. https://doi.org/10.2307/747641

Nie, N. H. (2001). Sociability, interpersonal relations, and the Internet. *American Behavioral Scientist, 45*(3), 420–435. https://doi.org/10.1177/00027640121957277

Oberst, U., Wegmann, E., Stodt, B., Brand, M., & Chamarro, A. (2017). Negative consequences from heavy social networking in adolescents: The mediating role of fear of missing out. *Journal of Adolescence, 55*(1), 51–60. https://doi.org/10.1016/j.adolescence.2016.12.008

O'Donnell, S., & Epstein, L. H. (2019). Smartphones are more reinforcing than food for students. *Addictive Behaviors, 90*, 124–133. https://doi.org/10.1016/j.addbeh.2018.10.018

Orben, A. (2020). Teenagers, screens and social media: A narrative review of reviews and key studies. *Social Psychiatry and Psychiatric Epidemiology, 55*(4), 407–414. https://doi.org/10.1007/s00127-019-01825-4

Parry, D. A., Davidson, B. I., Sewall, C. J. R., Fisher, J. T., Mieczkowski, H., & Quintana, D. S. (2021). A systematic review and meta-analysis of discrepancies between logged and self-reported digital media use. *Nature Human Behaviour, 5*(11), 1535–1547. https://doi.org/10.1038/s41562-021-01117-5

Paus, T., Keshavan, M., & Giedd, J. N. (2008). Why do many psychiatric disorders emerge during adolescence? Nature reviews. *Neuroscience, 9*(12), 947–957. https://doi.org/10.1038/nrn2513

Pera, A. (2018). Psychopathological processes involved in social comparison, depression, and envy on Facebook. *Frontiers in Psychology, 9*, Article 22. https://doi.org/10.3389/fpsyg.2018.00022

Plummer, F., Manea, L., Trepel, D., & McMillan, D. (2016). Screening for anxiety disorders with the GAD-7 and GAD-2: A systematic review and diagnostic meta-analysis. *General Hospital Psychiatry, 39*, 24–31. https://doi.org/10.1016/j.genhosppsych.2015.11.005

Przybylski, A. K., Murayama, K., DeHaan, C. R., & Gladwell, V. (2013). Motivational, emotional, and behavioral correlates of fear of missing out. *Computers in Human Behavior, 29*(4), 1841–1848. https://doi.org/10.1016/j.chb.2013.02.014

Radloff, L. S. (1977). The CES-D Scale. *Applied Psychological Measurement, 1*(3), 385–401. https://doi.org/10.1177/014662167700100306

Roberts, J., & Koliska, M. (2014). The effects of ambient media: What unplugging reveals about being plugged in. *First Monday, 19*(8). https://doi.org/10.5210/fm.v19i8.5220

Ross, R., Chaput, J.-P., Giangregorio, L. M., Janssen, I., Saunders, T. J., Kho, M. E., Poitras, V. J., Tomasone, J. R., El-Kotob, R., McLaughlin, E. C., Duggan, M., Carrier, J., Carson, V., Chastin, S. F., Latimer-Cheung, A. E., Chulak-Bozzer, T., Faulkner, G., Flood, S. M., Gazendam, M. K., . . . Tremblay, M. S. (2020). Canadian 24-hour movement guidelines for adults aged 18–64 years and adults aged 65 years or older: An integration of physical activity, sedentary behaviour, and sleep. *Applied Physiology, Nutrition, and Metabolism, 45*(10 Suppl. 2), S57–S102. https://doi.org/10.1139/apnm-2020-0467

Sampasa-Kanyinga, H., & Chaput, J.-P. (2016). Use of social networking sites and adherence to physical activity and screen time recommendations in adolescents. *Journal of Physical Activity & Health, 13*(5), 474–480. https://doi.org/10.1123/jpah.2015-0340

Sampasa-Kanyinga, H., Chaput, J.-P., & Hamilton, H. A. (2019). Social media use, school connectedness, and academic performance among adolescents. *The Journal of Primary Prevention, 40*(2), 189–211. https://doi.org/10.1007/s10935-019-00543-6

Sampasa-Kanyinga, H., Goldfield, G. S., Kingsbury, M., Clayborne, Z., & Colman, I. (2020). Social media use and parent–child relationship: A

cross-sectional study of adolescents. *Journal of Community Psychology*, *48*(3), 793–803. https://doi.org/10.1002/jcop.22293

Sampasa-Kanyinga, H., & Hamilton, H. A. (2015). Social networking sites and mental health problems in adolescents: The mediating role of cyberbullying victimization. *European Psychiatry*, *30*(8), 1021–1027. https://doi.org/10.1016/j.eurpsy.2015.09.011

Sampasa-Kanyinga, H., Hamilton, H. A., & Chaput, J.-P. (2018). Use of social media is associated with short sleep duration in a dose-response manner in students aged 11 to 20 years. *Acta Paediatrica*, *107*(4), 694–700. https://doi.org/10.1111/apa.14210

Sampasa-Kanyinga, H., Lalande, K., & Colman, I. (2020). Cyberbullying victimisation and internalising and externalising problems among adolescents: The moderating role of parent-child relationship and child's sex. *Epidemiology and Psychiatric Sciences*, *29*, Article e8. https://doi.org/10.1017/S2045796018000653

Sariyska, R., Rathner, E.-M., Baumeister, H., & Montag, C. (2018). Feasibility of linking molecular genetic markers to real-world social network size tracked on smartphones. *Frontiers in Neuroscience*, *12*, Article 945. https://doi.org/10.3389/fnins.2018.00945

Scott, H., Biello, S. M., & Woods, H. C. (2019). Social media use and adolescent sleep patterns: Cross-sectional findings from the UK millennium cohort study. *BMJ Open*, *9*(9), Article e031161. https://doi.org/10.1136/bmjopen-2019-031161

Shrout, P. E., Stadler, G., Lane, S. P., McClure, M. J., Jackson, G. L., Clavél, F. D., Iida, M., Gleason, M. E. J., Xu, J. H., & Bolger, N. (2018). Initial elevation bias in subjective reports. *Proceedings of the National Academy of Sciences*, *115*(1), E15–E23. https://doi.org/10.1073/pnas.1712277115

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine*, *166*(10), 1092–1097. https://doi.org/10.1001/archinte.166.10.1092

Statistics Canada. (2019). *Physical activity and screen time among Canadian children and youth, 2016 and 2017*.

Statistics Canada. (2021, September 27). Survey on COVID-19 and mental health February to May 2021. *The Daily*. https://www150.statcan.gc.ca/n1/en/daily-quotidien/210927/dq210927a-eng.pdf?st=wSNSXM

Steers, M.-L. N., Wickham, R. E., & Acitelli, L. K. (2014). Seeing everyone else's highlight reels: How Facebook usage is linked to depressive symptoms. *Journal of Social and Clinical Psychology*, *33*(8), 701–731. https://doi.org/10.1521/jscp.2014.33.8.701

Steger, M. F., & Kashdan, T. B. (2009). Depression and everyday social activity, belonging, and well-being. *Journal of Counseling Psychology*, *56*(2), 289–300. https://doi.org/10.1037/a0015416

Substance Abuse and Mental Health Services Administration. (2017). *Results from the 2017 National survey on drug use and health*.

Thai, H., Davis, C. G., Stewart, N., Gunnell, K. E., & Goldfield, G. S. (2021). The effects of reducing social media use on body esteem among transitional-aged youth. *Journal of Social and Clinical Psychology*, *40*(6), 481–507. https://doi.org/10.1521/jscp.2021.40.6.481

Thorisdottir, I. E., Sigurvinsdottir, R., Asgeirsdottir, B. B., Allegrante, J. P., & Sigfusdottir, I. D. (2019). Active and passive social media use and symptoms of anxiety and depressed mood among Icelandic adolescents. *Cyberpsychology, Behavior, and Social Networking*, *22*(8), 535–542. https://doi.org/10.1089/cyber.2019.0079

Tiirikainen, K., Haravuori, H., Ranta, K., Kaltiala-Heino, R., & Marttunen, M. (2019). Psychometric properties of the 7-item Generalized Anxiety Disorder Scale (GAD-7) in a large representative sample of Finnish adolescents. *Psychiatry Research*, *272*, 30–35. https://doi.org/10.1016/j.psychres.2018.12.004

Tromholt, M. (2016). The Facebook experiment: Quitting Facebook leads to higher levels of well-being. *Cyberpsychology, Behavior and Social Networking*, *19*(11), 661–666. https://doi.org/10.1089/cyber.2016.0259

Turel, O. (2015). Quitting the use of a habituated hedonic information system: A theoretical model and empirical examination of Facebook users. *European Journal of Information Systems*, *24*(4), 431–446. https://doi.org/10.1057/ejis.2014.19

Turel, O., Cavagnaro, D. R., & Meshi, D. (2018). Short abstinence from online social networking sites reduces perceived stress, especially in excessive users. *Psychiatry Research*, *270*, 947–953. https://doi.org/10.1016/j.psychres.2018.11.017

Twenge, J. M., & Campbell, W. K. (2019). Media use is linked to lower psychological well-being: Evidence from three datasets. *The Psychiatric Quarterly*, *90*(2), 311–331. https://doi.org/10.1007/s11126-019-09630-7

Verduyn, P., Ybarra, O., Résibois, M., Jonides, J., & Kross, E. (2017). Do social network sites enhance or undermine subjective well-being? A critical review. *Social Issues and Policy Review*, *11*(1), 274–302. https://doi.org/10.1111/sipr.12033

Walther, J. B., Liang, Y. J., DeAndrea, D. C., Tong, S. T., Carr, C. T., Spottswood, E. L., & Amichai-Hamburger, Y. (2011). The effect of feedback on identity shift in computer-mediated communication. *Media Psychology*, *14*(1), 1–26. https://doi.org/10.1080/15213269.2010.547832

Wegmann, E., Oberst, U., Stodt, B., & Brand, M. (2017). Online-specific fear of missing out and internet-use expectancies contribute to symptoms of internet-communication disorder. *Addictive Behaviors Reports*, *5*, 33–42. https://doi.org/10.1016/j.abrep.2017.04.001

Woods, H. C., & Scott, H. (2016). #Sleepyteens: Social media use in adolescence is associated with poor sleep quality, anxiety, depression and low self-esteem. *Journal of Adolescence*, *51*(1), 41–49. https://doi.org/10.1016/j.adolescence.2016.05.008

Received May 10, 2023
Revision received January 10, 2024
Accepted February 26, 2024 ■

Psychology of Popular Media

ISSN: 2689-6567

https://doi.org/10.1037/ppm0000560

Correction to "Limiting Social Media Use Decreases Depression, Anxiety, and Fear of Missing Out in Youth With Emotional Distress: A Randomized Controlled Trial" by Davis and Goldfield (2024)

In the article "Limiting Social Media Use Decreases Depression, Anxiety, and Fear of Missing Out in Youth With Emotional Distress: A Randomized Controlled Trial" by Christopher G. Davis and Gary S. Goldfield (*Psychology of Popular Media*, 2025, Vol. 14, No. 1, pp. 1–11. https://doi.org/10.1037/ppm0000536), the mean and standard deviation in Table 1 for the sleep (hour/day) variable for the intervention and control conditions should instead appear as 8.27 (1.49) and 8.43 (1.27), respectively. All versions of this article have been corrected.

https://doi.org/10.1037/ppm0000560

American Economic Review 2022, 112(11): 3660–3693
https://doi.org/10.1257/aer.20211218

Social Media and Mental Health†

By Luca Braghieri, Ro'ee Levy, and Alexey Makarin*

Quasi-experiment

Self-report

...al estimates of the impact of social ...eraging a unique natural experiment: ...' Facebook across US colleges. Our ...nt mental health around the years of ... generalized difference-in-differences ...t the rollout of Facebook at a college ...nt mental health. It also increased the ...: reported experiencing impairments ...o poor mental health. Additional evi... the results are due to Facebook fos...arisons. (*JEL* D91, I12, I23, L82)

...an half of the world population—had a social ...ent around two and a half hours per day on ...'e Are Social 2021). Very few technologies ... shaped the way people spend their time and interact with others.

As social media started gaining popularity in the mid-2000s, the mental health of adolescents and young adults in the United States began to worsen (Patel et al. 2007; Twenge et al. 2019). For instance, the total number of individuals aged 18–23

*Braghieri: Bocconi University, IGIER, CEPR, and CESifo (email: luca.braghieri@unibocconi.it); Levy: Tel Aviv University (email: roeelevy@tauex.tau.ac.il); Makarin: MIT Sloan School of Management, EIEF, and CEPR (email: makarin@mit.edu). Stefano DellaVigna was the coeditor for this article. We would like to thank Sarah Eichmeyer for her contributions at the early stages of this project, Mary Hoban for helping us access the NCHA dataset, and Luis Armona for his collaboration in putting together the Facebook expansion dates dataset. We are grateful to the editor, three anonymous referees, Levi Boxell, Davide Cantoni, Daniel Chen, Scott Cunningham, Georgy Egorov, Ruben Enikolopov, Amy Finkelstein, Thomas Fujiwara, Matthew Gentzkow, Luigi Guiso, Martin Mattsson, Jack Mountjoy, Samuel Norris, Petra Persson, Maria Petrova, Andrea Prat, Maya Rossin-Slater, Frank Schilbach, Sebastian Schweighofer-Kodritsch, Joseph Shapiro, Andrey Simonov, and seminar participants at EIEF, LMU, Max Planck Institute, MIT, NHH Norwegian School of Economics, QMUL, TSE, University of Frankfurt, UPF, the Italian National Health Institute (Istituto Superiore di Sanità), the UK Office of Communications (Ofcom), and conference participants at the Center for Rationality and Competition, CEPR Political Economy and Populism Symposium, CEPR-EIEF-Tor Vergata Media Workshop, the Early Career Behavioral Economics Conference (ECBE), NBER IT & Digitization, and the Southern Economic Association for helpful comments. Luca Braghieri gratefully acknowledges financial support from the Deutsche Forschungsgemeinschaft through CRC TRR 190 (project 280092119). Ro'ee Levy gratefully acknowledges financial support from the Foerder Institute for Economic Research. We thank Beikeren Büyükeren, Juan Carlos Cisneros, Valerio Sergio Castaldo, Gleb Kozlyakov, Oksana Kuznetsova, and Merujert Tatkeyeva for excellent research assistance. The opinions, findings, and conclusions reported in this article are those of the authors and are in no way meant to represent the corporate opinions, views, or policies of the American College Health Association (ACHA). ACHA does not warrant nor assume any liability or responsibility for the accuracy, completeness, or usefulness of any information presented in this article. Disclosures: Levy is an unpaid member of Facebook's 2020 Election Research Project.

†Go to https://doi.org/10.1257/aer.20211218 to visit the article page for additional materials and author disclosure statements.

American Economic Review 2022, 112(11): 3660–3693
https://doi.org/10.1257/aer.20211218

Social Media and Mental Health†

By Luca Braghieri, Ro'ee Levy, and Alexey Makarin*

We provide quasi-experimental estimates of the impact of social media on mental health by leveraging a unique natural experiment: the staggered introduction of Facebook across US colleges. Our analysis couples data on student mental health around the years of Facebook's expansion with a generalized difference-in-differences empirical strategy. We find that the rollout of Facebook at a college had a negative impact on student mental health. It also increased the likelihood with which students reported experiencing impairments to academic performance due to poor mental health. Additional evidence on mechanisms suggests the results are due to Facebook fostering unfavorable social comparisons. (*JEL* D91, I12, I23, L82)

In 2021, 4.3 billion people—more than half of the world population—had a social media account, and the average user spent around two and a half hours per day on social media platforms (GWI 2021; We Are Social 2021). Very few technologies since television have so dramatically reshaped the way people spend their time and interact with others.

As social media started gaining popularity in the mid-2000s, the mental health of adolescents and young adults in the United States began to worsen (Patel et al. 2007; Twenge et al. 2019). For instance, the total number of individuals aged 18–23

*Braghieri: Bocconi University, IGIER, CEPR, and CESifo (email: luca.braghieri@unibocconi.it); Levy: Tel Aviv University (email: roeelevy@tauex.tau.ac.il); Makarin: MIT Sloan School of Management, EIEF, and CEPR (email: makarin@mit.edu). Stefano DellaVigna was the coeditor for this article. We would like to thank Sarah Eichmeyer for her contributions at the early stages of this project, Mary Hoban for helping us access the NCHA dataset, and Luis Armona for his collaboration in putting together the Facebook expansion dates dataset. We are grateful to the editor, three anonymous referees, Levi Boxell, Davide Cantoni, Daniel Chen, Scott Cunningham, Georgy Egorov, Ruben Enikolopov, Amy Finkelstein, Thomas Fujiwara, Matthew Gentzkow, Luigi Guiso, Martin Mattson, Jack Mountjoy, Samuel Norris, Petra Persson, Maria Petrova, Andrea Prat, Maya Rossin-Slater, Frank Schilbach, Sebastian Schweighofer-Kodritsch, Joseph Shapiro, Andrey Simonov, and seminar participants at EIEF, LMU, Max Planck Institute, MIT, NHH Norwegian School of Economics, QMUL, TSE, University of Frankfurt, UPF, the Italian National Health Institute (Istituto Superiore di Sanità), the UK Office of Communications (Ofcom), and conference participants at the Center for Rationality and Competition, CEPR Political Economy and Populism Symposium, CEPR-EIEF-Tor Vergata Media Workshop, the Early Career Behavioral Economics Conference (ECBE), NBER IT & Digitization, and the Southern Economic Association for helpful comments. Luca Braghieri gratefully acknowledges financial support from the Deutsche Forschungsgemeinschaft through CRC TRR 190 (project 280092119). Ro'ee Levy gratefully acknowledges financial support from the Foerder Institute for Economic Research. We thank Berkeren Büyükeren, Juan Carlos Cisneros, Valerio Sergio Castaldo, Gleb Kozlyakov, Oksana Kuznetsova, and Meruyert Tatkeyeva for excellent research assistance. The opinions, findings, and conclusions reported in this article are those of the authors and are in no way meant to represent the corporate opinions, views, or policies of the American College Health Association (ACHA). ACHA does not warrant nor assume any liability or responsibility for the accuracy, completeness, or usefulness of any information presented in this article. Disclosures: Levy is an unpaid member of Facebook's 2020 Election Research Project.

†Go to https://doi.org/10.1257/aer.20211218 to visit the article page for additional materials and author disclosure statements.

who reported experiencing a major depressive episode in the past year increased by 83 percent between 2008 and 2018 (NSDUH 2019). Similarly, over the same time period, suicides became more prevalent and are now the second leading cause of death for individuals 15–24 years old (National Center for Health Statistics 2021). Although the ultimate causes of these trends are still largely unknown, scholars have hypothesized that the diffusion of social media might be an important contributing factor (Twenge et al. 2019). In fact, concerns about a potential negative effect of social media on mental health have become so prominent that the US Senate held a committee hearing about the topic in late 2021. Well-identified causal evidence, however, remains scarce.

In this paper, we provide quasi-experimental estimates of the impact of social media on mental health by leveraging a unique natural experiment: the staggered introduction of Facebook across US colleges in the mid-2000s. Coupling survey data on college students' mental health collected in the years around Facebook's expansion with a generalized difference-in-differences empirical strategy, we find that the introduction of Facebook at a college had a negative impact on student mental health. We also find that, after the introduction of Facebook, students were more likely to report that poor mental health negatively affected their academic performance. Finally, we present an array of additional evidence suggesting that the results are consistent with Facebook enhancing students' abilities to engage in unfavorable social comparisons.

The early expansion of Facebook across colleges in the United States is a particularly promising setting to investigate the effects of social media use on the mental health of young adults. Facebook was created at Harvard in February 2004, but it was only made available to the general public in September 2006. Between February 2004 and September 2006, Facebook was rolled out across US colleges in a staggered fashion. Upon being granted access to Facebook's network, colleges witnessed rapid and widespread Facebook penetration among students (Wilson, Gosling, and Graham 2012; Brügger 2015). The staggered and sharp introduction of Facebook across US colleges provides a source of quasi-experimental variation in exposure to social media that we can leverage for causal identification.

We employ two main datasets in our analysis: the first dataset specifies the dates in which Facebook was introduced at 775 US colleges; the second consists of the universe of answers to seventeen consecutive waves of the National College Health Assessment (NCHA), the most comprehensive survey about student mental and physical health available at the time of Facebook's expansion.

Our analysis relies on a generalized difference-in-differences research design, where one of the dimensions of variation is the college a student attends, and the other dimension is whether the student took the survey before or after the introduction of Facebook at her college. Under a parallel trends assumption, the college by survey-wave variation generated by the sharp but staggered introduction of Facebook allows us to obtain causal estimates of the introduction of Facebook on student mental health. Our empirical strategy allows us to rule out various confounding factors: first, college-specific differences fixed in time (e.g., students at more academically demanding colleges may have worse baseline mental health than students at less demanding colleges); second, differences across time that affect all students in a similar way (e.g., certain macroeconomic fluctuations); third, mental

health trends affecting colleges in different Facebook expansion groups differentially, but smoothly (e.g., colleges where Facebook was rolled out earlier may be on different linear trends in terms of mental health than colleges where Facebook was rolled out later).[1] We also address recent econometric concerns with staggered difference-in-differences research designs by showing robustness to the use of a variety of alternative estimators.[2] Lastly, we complement the difference-in-differences strategy with a specification that exploits variation in length of exposure to Facebook across students within a college and survey wave, and that, therefore, does not rely on our baseline college-level parallel trends assumption for identification.

Our main finding is that the introduction of Facebook at a college had a negative effect on student mental health. Our index of poor mental health, which aggregates all the relevant mental health variables in the NCHA survey, increased by 0.085 standard deviation units as a result of the introduction of Facebook. As a point of comparison, this magnitude is around 22 percent of the effect of losing one's job on mental health, as reported in a meta-analysis by Paul and Moser (2009). We further benchmark our results against a clinically validated depression scale: the Patient Health Questionnaire-9 (PHQ-9; Kroenke, Spitzer, and Williams 2001). The effect of the introduction of Facebook on our index of poor mental health is equivalent to a 2 percentage point increase in the share of students suffering from depression according to the PHQ-9 over a baseline of 25 percent. Lastly, we perform a back-of-the-envelope calculation to determine what fraction of the increased prevalence of severe depression among college students over the last two decades can be explained by the introduction of Facebook. Under a set of relatively strong assumptions, we calculate that the introduction of Facebook accounts for approximately 24 percent of such increase.

We highlight three additional results. First, the negative effects on mental health are strongest for students who, based on immutable characteristics such as gender and age, are predicted to be most susceptible to mental illness. For those students, we also observe a significant increase in depression diagnoses, take-up of psychotherapy for depression, and use of antidepressants. Second, in the short-to-medium run, the negative effects of Facebook on mental health increase with length of exposure to the platform. Third, students reported suffering some negative downstream effects as a result of their worsened mental health conditions. Specifically, after the introduction of Facebook, students were more likely to report that their academic performance was negatively affected by conditions related to poor mental health.

What explains the negative effects of Facebook on mental health? The pattern of results is consistent with Facebook increasing students' ability to engage in unfavorable social comparisons. Two main pieces of evidence bear on this conclusion. First, we find that the effects are particularly pronounced for students who might view themselves as comparing unfavorably to their peers, such as students who live off-campus—and therefore are more likely to be excluded from on-campus social activities—students of lower socioeconomic status, and students not belonging to fraternities/sororities. Second, we show that the introduction of Facebook directly

[1]The last confounding factor in the list is taken into account in a specification that includes linear time trends at the Facebook-expansion-group level.

[2]See Roth et al. (2022) for a recent overview.

affected the students' beliefs about their peers' social lives and behaviors. Consistent with the content on Facebook at the time, changes in perceptions are limited to alcohol as opposed to other drugs. As far as other channels are concerned, we do not find significant evidence that the negative effects of Facebook on mental health are due to disruptive internet use. We also rule out several alternative mechanisms, such as reduced stigma about mental illness and direct effects on drug use, alcohol consumption, and sexual behaviors.

The results presented in this paper, which rely on the staggered rollout of Facebook across US college campuses in 2004–2006, should be interpreted with caution for several reasons. First, our estimates cannot speak directly to the effects of social media features (e.g., news pages) that were introduced after the time period we analyze. Similarly, our estimates cannot speak directly to the possibility that years of experience with the platform might teach users ways to mitigate the negative effects on mental health.[3] Second, despite being the core component of most mental health diagnoses, self-reports may still suffer from measurement error due to recall bias, lack of incentives, and social image concerns.[4] Finally, we note that our estimates are local to college students, a population of direct interest in the discussion about the recent worsening of mental health trends among adolescents and young adults. Nevertheless, future research should test whether social media has a similar effect on the mental health of other demographic groups.

Aside from these caveats, our findings are in line with the hypothesis that social media has a negative impact on mental health and played a role in the increase in mental illness among adolescents and young adults over the past two decades. Of course, our results do not imply that the overall welfare effects of social media are necessarily negative. Such calculation would require estimating the effects of social media along various other dimensions; furthermore, they would require taking into account potential positive effects, such as a reduction in the cost of connecting with friends and family across a distance. Nevertheless, we believe our results will be informative to social media users and policymakers alike.

This paper contributes to the literature by providing the most comprehensive causal evidence to date on the effects of social media on mental health. The three closest papers to ours (Allcott et al. 2020; Mosquera et al. 2020; and Allcott, Gentzkow, and Song 2021), feature experiments that incentivize a randomly selected subset of participants to reduce their social media use.[5] Those studies find negative effects of social media use on self-reported well-being, and Allcott, Gentzkow, and Song (2021) shows evidence of digital addiction. Our findings complement the aforementioned literature in five main ways. First, our mental health outcome variables are more comprehensive and detailed than the ones in the experimental papers above. Specifically, our outcome variables include 11 questions related to depression (covering symptoms, diagnoses, take-up of psychotherapy, and use of antidepressants)

[3] One of our specifications, equation (4), can look at up to two and a half years of experience with the platform and finds that the effects, if anything, increase in the short to medium term. Longer-term effects, however, could be quite different.

[4] The effects on academic performance are also self-reported and could suffer from similar issues.

[5] For correlational evidence on the link between social media and mental health, see Lin et al. (2016); Dienlin, Masur, and Trepte (2017); Berryman, Ferguson, and Negy (2018); Kelly et al. (2019); Bekalu, McCloud, and Viswanath (2019); Twenge and Campbell (2019).

and various questions related to other mental health conditions, ranging from seasonal affective disorder to anorexia. By contrast, the three experimental studies above measure broadly defined subjective well-being and include only one question that relates directly to a mental health condition listed in the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5). Second, rather than studying the partial equilibrium effects of paying isolated individuals to reduce their social media use, our estimates capture the general equilibrium effects of introducing social media in an entire community.[6] Such general equilibrium effects are arguably particularly important for technologies like social media that exhibit strong network externalities. Third, our analysis is less prone to experimenter demand, Hawthorne, and income effects.[7] Fourth, the experiments above study fairly short-term disruptions in social media use, ranging from 1 to 12 weeks; conversely, we can estimate effects up to several semesters after the introduction of Facebook at a college. Fifth, our study specifically targets the population (young adults) that experienced the most severe deterioration in mental health in recent decades and studies it around the time in which those mental health trends began to worsen. Focusing on young adults is arguably important for two additional reasons: first, because early adulthood may be a particularly vulnerable time as far as mental health is concerned (Kessler et al. 2007); second, because early adulthood is an age in which individuals often make critical life decisions.

This paper also relates to the rapidly growing literature in economics about the determinants and consequences of mental illness (Ridley et al. 2020). Research on the determinants of mental illness showed that unconditional cash transfers, in utero exposure to the death of a maternal relative, unemployment shocks, and economic downturns can affect mental health (Paul and Moser 2009; Haushofer and Shapiro 2016; Persson and Rossin-Slater 2018; Golberstein, Gonzales, and Meara 2019). Donati et al. (2022) provide quasi-experimental evidence that access to high-speed internet increased the incidence of mental disorders among young adults in Italy. We contribute to this strand of the literature by focusing on social media, which many consider to be an important driver of the recent rise in depression rates among adolescents and young adults (Twenge 2017; Twenge et al. 2019). Studies focusing on the consequences of mental illness found that better mental health is associated with fewer crimes, increased parental investment in children, and better labor market outcomes (Blattman, Jamison, and Sheridan 2017; Biasi, Dahl, and Moser 2019; Baranov et al. 2020; Shapiro 2021). We complement this literature by showing that, after the introduction of Facebook, students were more likely to report experiencing impairments to academic performance as a result of poor mental health.

[6] There are likely substantial endogenous adjustments of one's social media use to one's peers' social media use, as well as spillover effects on one's mental health due to one's peers' social media use. We employ the term "general equilibrium effects" to indicate that our estimates capture such indirect effects, as well as any direct effects.

[7] In the case of the experiments listed above, subjects in the treatment group are paid to reduce their social media use and are therefore not blind to treatment status, which might give rise to experimenter demand effects. In addition, the mere fact of being observed (e.g., via daily text messages asking participants how they feel) might affect subjects' behaviors independently of treatment status, giving rise to general Hawthorne effects. Lastly, incentive payments might directly affect self-reported well-being and confound interpretation. An additional issue with social media experiments is that they often screen participants who do not meet certain criteria and, therefore, employ rather selected samples. For instance, the main sample analyzed in Allcott et al. (2020) includes participants who reported using Facebook more than 15 minutes per day and were willing to accept $102 to deactivate their Facebook accounts for a month.

Lastly, this paper contributes to an emerging literature exploiting the expansion of social media platforms to study the effects of social media on a variety of outcomes. The empirical strategy adopted in this paper is closely related to the one in Armona (2019), who leverages the staggered introduction of Facebook across US colleges to study labor market outcomes more than a decade later. Enikolopov, Makarin, and Petrova (2020) and Fergusson and Molina (2020) exploit the expansion of the social media platform VK in Russia and of Facebook worldwide, respectively, to show that social media use increases protest participation. Bursztyn et al. (2019) and Müller and Schwarz (2020) exploit the expansion of VK and Twitter, respectively, and find that social media use increases the prevalence of hate crimes.[8] A unique feature of our setting is that it allows us to measure the effects of the sharp rollout of the biggest social media platform in the world at a time in which very few close substitutes were available.

The remainder of the paper is organized as follows: Section I provides some background on mental health and on Facebook's early expansion; Section II describes the data sources used in the analysis and presents descriptive statistics; Section III discusses the empirical strategy; Section IV presents the results; Section V explores mechanisms; Section VI discusses potential implications of the results; Section VII concludes.

I. Background

Mental Health.—Mental illnesses, such as depression, anxiety, bipolar disorder, and schizophrenia, are disturbingly common and can be highly debilitating. According to the Global Burden of Disease study, around a billion people in the world suffered from mental disorders in 2017, with depression and anxiety-related disorders as the leading conditions (James et al. 2018). In the United States, around 1 in 5 adults experiences some form of mental illness each year, and 1 in 20 experiences serious mental illness (NAMI 2020). Mental health conditions can have severe adverse effects, hampering people's ability to work, study, and be productive. According to the World Health Organization's Global Burden of Disease, mental illness is the most burdensome disease category in terms of total disability-adjusted years for adults younger than 45, and depression is one of the most taxing conditions (World Health Organization 2008; Layard 2017).

Alarmingly, the last two decades witnessed a worsening of mental health trends in the United States, especially among adolescents and young adults (Twenge et al. 2019). As shown in online Appendix Figure A.1, self-reported episodes of psychological distress and depression have risen substantially over the past 15 years, with the highest growth rate among young adults. Similarly, both self-reports of suicidal thoughts, plans, or attempts and actual suicides have increased considerably among that demographic group. Since the timing of the divergence in mental health trends between young adults and older generations roughly coincides with wider adoption

[8] Additional research on social media and political outcomes includes Enikolopov, Petrova, and Sonin (2018); Fujiwara, Müller, and Schwarz (2021); and Levy (2021). For a detailed overview, see Zhuravskaya, Petrova, and Enikolopov (2020).

of social media, various scholars have hypothesized the two phenomena might be related (Twenge 2017; Twenge et al. 2019).

A Brief History of Facebook's Expansion and Initial Popularity.—Facebook was created at Harvard in February 2004 and was rolled out gradually to other colleges in the United States and abroad over the subsequent two and a half years. The staggered nature of the rollout was enforced by requiring users to be in possession of verified email addresses (e.g., addresses ending in @harvard.edu). The rollout of Facebook across US colleges was not random: as shown in the descriptive statistics section, more selective colleges were granted access to Facebook relatively earlier than less selective colleges. The staggered nature of the expansion is arguably due to three factors: first, scale constraints due to limited server capacity (Kirkpatrick 2011); second, Facebook's willingness, at least at the outset, to maintain a flavor of exclusivity; third, Facebook's desire to strengthen network effects by keeping the fraction of users who likely knew each other offline artificially high (Aral 2021).

Even in its infancy, Facebook was extremely popular. Upon being granted access to the platform, colleges witnessed rapid and very widespread adoption among students.[9] To get a sense of the early adoption rates among college students, we matched data provided by Facebook on the number of users at each of the first 100 colleges that were granted access to the platform with IPEDS (Integrated Postsecondary Education Data System) data on the number of full-time undergraduate students at those colleges (US Department of Education 2005; Traud, Mucha, and Porter 2012). Online Appendix Figure A.2 presents a histogram of the number of users per 100 undergraduate students at those colleges and shows that, in September 2005, there were on average 86 Facebook users for every 100 undergraduate students. This result is consistent with Facebook's statement that, across all the colleges with access to the platform as of September 2005, approximately 85 percent of students had a Facebook profile (Arrington 2005).[10]

Not only was Facebook immediately popular, usage was also quite intense. In early 2006, close to three-quarters of users logged into the site at least once a day, and the average user logged in six times a day (Hass 2006). As of early 2006, Facebook was the ninth most visited website on the internet, despite not yet being open to the general public (Hass 2006).

II. Data Sources and Descriptive Statistics

A. *Data Sources*

Our analysis relies primarily on two data sources. The first data source specifies the dates in which Facebook was introduced at 775 US colleges. The second

[9] According to a description by Kirkpatrick (2011, p. 88), "within days, [Facebook] typically captured essentially the entire student body, and more than 80 percent of users returned to the site daily."

[10] Various smaller-scale studies using survey and/or Facebook data and focusing mostly on undergraduate students confirm the high adoption rates in 2005–2006. Specifically, those studies show that, at the colleges in which they were administered, 82–94 percent of students had a Facebook account (Stutzman 2006; Kolek and Saunders 2008; Lampe, Ellison, and Steinfield 2008). While women may have been more likely than men to join Facebook, Facebook usage was very common across demographic groups (Kolek and Saunders 2008).

consists of the universe of answers to seventeen consecutive waves of the NCHA survey, the largest and most comprehensive survey on college students' mental and physical health at the time of Facebook's expansion.

Facebook Expansion Dates Data.—The Facebook expansion dates dataset was assembled in two steps: for the first 100 colleges that received Facebook access, we rely on introduction dates collected and made public in previous studies (Traud, Mucha, and Porter 2012; Jacobs et al. 2015). For the remaining 675 colleges in the dataset, we obtained Facebook introduction dates using the Wayback Machine, an online archive that contains snapshots of various websites at different points in time and that allows users to visit old versions of those websites. Specifically, between the spring of 2004 and the spring of 2005, the front page of Facebook's website was regularly updated to show the most recent set of colleges that had been given access to the platform.[11] As an example, online Appendix Figure A.3 shows the front page of Facebook as of June 15 2004, recovered via the Wayback Machine. As shown in the figure, Facebook was open to 34 colleges at that point in time.

Armed with a time series of snapshots of the front page of Facebook's website, it is possible to reconstruct tentative dates in which Facebook was rolled out at each college. Specifically, the rollout date at a certain college should be between the date of the first snapshot in which the college is listed and the date of the previous snapshot. When the distance between the snapshots is more than one day, we consider the first date in which a college is listed on Facebook's front page as the introduction date.

Since the Wayback Machine took snapshots of Facebook's Website at a high temporal resolution, our imputation procedure for the introduction dates is rather precise. For the months in which our introduction dates rely on the Wayback Machine (September 2004 to May 2005) the average number of days between consecutive snapshots is one and a half. Therefore, on average, our imputed introduction dates should be within two days of the actual introduction dates.

Online Appendix Table A.32 lists the colleges in the Facebook expansion dates dataset and the date in which Facebook was rolled out at each of them.

NCHA Data.—Our second main data source consists of more than 430,000 responses to the NCHA survey, a survey administered to college students on a semi-annual basis by the American College Health Association (ACHA). The NCHA survey was developed in 1998 by a team of college health professionals with the purpose of obtaining information from college students about their mental and physical health. Specifically, the NCHA survey inquires about demographics, physical health, mental health, alcohol and drug use, sexual behaviors, and perceptions of these behaviors among one's peers.

As far as mental health is concerned, the NCHA survey includes both questions about symptoms of mental illness and questions about take-up of mental healthcare services. We emphasize that reliance on self-reported symptoms is part of standard

[11] Beginning with the fall of 2005, Facebook started listing the colleges that had access to the platform on a separate page that is snapshotted too infrequently to allow us to extract meaningful introduction dates. Therefore, our Facebook introduction dates dataset ends after the spring of 2005.

medical practice in the domain of mental health (Chan 2010). Specifically, according to the official diagnostic manual of the American Psychiatric Association (DSM-5), the diagnosis of many mental health disorders including depression relies almost exclusively on patients' self-reports of symptoms such as difficulty sleeping, anhedonia, fatigue, feelings of worthlessness and guilt, diminished ability to think or concentrate, and recurrent suicidal thoughts (APA 2013). In fact, self-administered questionnaires inquiring about depression symptoms have been shown to predict medical diagnoses with accuracy rates up to 90 percent (Kroenke and Spitzer 2002).[12]

The NCHA dataset includes the universe of responses to all NCHA survey waves administered between the spring of 2000 and the spring of 2008, the longest stretch of time around Facebook's early expansion in which the content of the survey remained constant.[13] Colleges included in the NCHA dataset administered the survey to randomly selected classrooms, randomly selected students, or all students. The average response rate across the survey waves for which we have such information is 37 percent (ACHA 2000–2019). In order to assuage concerns about the possibility that the introduction of Facebook affected the composition of students who participated in the survey, online Appendix Tables A.3 and A.10 show that, along the demographic characteristics elicited in the NCHA survey, there are no meaningful compositional changes following the introduction of Facebook. Throughout our analysis, we limit our sample to full-time undergraduate students.[14]

The NCHA dataset is an unbalanced panel, in which colleges drop in and out. Specifically, every college in the United States can voluntarily select into any wave of the NCHA survey and is not required to keep administering the survey in subsequent years. To account for compositional changes to the panel, our preferred specification includes college fixed effects.

The NCHA survey does not include any questions on social media use; therefore, it is not possible for us to determine whether a particular survey respondent had a Facebook account. It is, however, possible to determine whether the college attended by the survey respondent had Facebook access at the time in which the respondent took the survey. In order to protect the privacy of the institutions that participate in the NCHA survey while still allowing us to carry out the analysis, the ACHA kindly agreed to provide us with a customized dataset that includes a variable indicating the semester in which Facebook was rolled out at each college. Specifically, the ACHA adopted the following procedure: (i) merge our dataset containing the Facebook introduction dates to the NCHA dataset; (ii) add a variable listing the semester in which Facebook was rolled out at each college;[15] (iii) strip

[12] Section III, online Appendix B, and online Appendix C discuss our symptom measures in detail and present an array of exercises to validate them.

[13] Between 1998 and 2000, the survey was being fine-tuned and changed considerably across survey waves; similarly, after the spring of 2008, the survey underwent a major revision that substantially limits comparability to previous waves.

[14] Graduate students also had access to the Facebook platform, but take-up was a lot smaller among graduate students than among undergraduates (e.g., Acquisti and Gross 2006).

[15] For the set of colleges that appear both in our introduction dates dataset and the NCHA survey, the ACHA listed the semesters corresponding to the introduction dates in our dataset. For the set of colleges that appear only in the NCHA dataset, we list the fall of 2005 as the semester in which Facebook was introduced at those colleges. Such imputation is sensible in virtue of the fact that our introduction dates dataset ends after the spring

away any information that could allow us to identify colleges (including the specific date in which Facebook was introduced at each college).

B. *Descriptive Statistics*

Online Appendix Tables A.1 and A.2 present college-level and student-level descriptive statistics for colleges in different Facebook expansion groups.[16] Online Appendix Table A.1 shows that colleges in earlier Facebook expansion groups are more selective in terms of test scores, larger, more likely to be on the East Coast, and have more residential undergraduate programs than colleges in later Facebook expansion groups. Panel A of online Appendix Table A.2, which averages student-level variables available in the NCHA dataset across the different Facebook expansion groups, shows that colleges in earlier Facebook expansion groups enroll students from relatively more advantaged economic backgrounds. Lastly, panel B of online Appendix Table A.2 shows that students in colleges that received Facebook relatively earlier have worse baseline mental health outcomes than students attending colleges in later Facebook expansion groups.[17] The baseline differences across Facebook expansion groups may lead one to wonder about the plausibility of the parallel trends assumption in this setting; we address concerns related to parallel trends in Section III.

Online Appendix Table A.1 also shows the number of colleges in the NCHA dataset that received Facebook access in each semester between the spring of 2004 and the fall of 2005. Other than the spring of 2004, when Facebook was first introduced, the fraction of colleges in the NCHA dataset that received Facebook access in each semester is fairly equally distributed across the remaining introduction semesters.

III. Empirical Strategy

Construction of the Primary Outcome Variables.—In order to mitigate concerns about cherry-picking outcome variables, we consider all the questions in the NCHA survey that are related to mental health and that inquire about a respondent's recent past (e.g., "Within the last school year, how many times have you felt so depressed that it was difficult to function?").

To impose structure on our analysis and assuage concerns about multiple hypothesis testing, we group the individual mental health variables into nested families and combine them into indices. The coarsest level of analysis combines all the mental health questions (*index of poor mental health*); a second level of analysis splits symptoms of mental illness (*index symptoms poor mental health*) and

semester of 2005 and that, by the end of 2005, the vast majority of US colleges had been granted access to Facebook. As shown in online Appendix A, the results are robust to dropping those colleges altogether.

[16] Online Appendix Table A.1 was constructed by merging our Facebook expansion dates dataset with data from IPEDS. We cannot directly provide college-level summary statistics using the NCHA dataset, because most college-level information in the NCHA was stripped away for privacy reasons.

[17] The differences in baseline mental health across Facebook expansion groups are particularly stark when comparing the first Facebook expansion group to the other groups; among the other groups the differences are more muted. In online Appendix A, we present and discuss a robustness check showing that our results do not significantly change when we drop colleges in each expansion group in turn or when we interact college-level baseline mental health with survey-wave fixed effects.

self-reported take-up of depression-related services (*index depression services*) into separate families; a third level of analysis splits the symptoms of mental illness into depression-related symptoms (*index of depression symptoms*) and symptoms related to other mental health conditions (*index symptoms other mental health conditions*); the finest level comprises the individual variables themselves.

The index of depression symptoms includes questions that inquire as to whether a student exhibited various symptoms of depression such as feeling hopeless, overwhelmed, exhausted, very sad, debilitatingly depressed, seriously considered committing suicide, or attempted suicide. The index of symptoms of other mental health conditions includes questions that inquire as to whether a student experienced issues related to anorexia, anxiety disorder, bulimia, and seasonal affective disorder. The overall index of symptoms of poor mental health encompasses both sets of symptoms.

The index of depression services requires a slightly more detailed discussion due to a peculiarity in the way the questions were structured. Specifically, the NCHA survey asked three questions about depression-related services: (i) whether the student was diagnosed with depression within the year prior to taking the survey, (ii) whether the student was in therapy for depression at the time in which she took the survey, and (iii) whether the student was on antidepressants at the time in which she took the survey. The NCHA survey asked those questions only to students who had given an affirmative answer to a previous question inquiring as to whether they had ever been diagnosed with depression. Therefore, the variables related to the three questions above should be interpreted as "having ever received a depression diagnosis" plus "having received a depression diagnosis in the last year", or "being in therapy for depression," or "taking antidepressants." Under this interpretation, we can safely impute zeros to the three questions about depression-related services for students who gave a negative answer to the question about whether they had ever been diagnosed with depression.

Our indices are constructed as follows: first, we orient all variables that compose an index in such a way that higher values always indicate worse mental health outcomes; second, we standardize those variables using means and standard deviations from the preperiod; third, we take an equally weighted average of the index components, excluding from the analysis observations in which any of the components are missing; fourth, we standardize the final index. This way, our indices are essentially z-scores.[18]

Online Appendix Table A.31 lists all the variables used in our analysis, describes their construction in detail, and includes the exact wording of the questions in the NCHA survey that each variable is based on.

Validation of the Primary Outcome Variables.—We validate the NCHA survey questions that form the basis of our primary outcome variables both internally and externally. We validate the questions about symptoms of mental illness internally by relating them to self-reported mental healthcare diagnoses within our dataset. Online Appendix B presents an array of validation exercises suggesting that the

[18] In online Appendix A, we show that are results are unchanged if we construct the indices in other ways, for instance as described in Anderson (2008).

questions about symptoms of mental illness in the NCHA survey are indeed highly predictive of mental illness diagnoses.

We validate the NCHA survey questions externally by conducting an original survey on more than 500 college students. Our survey contained both the questions from the NCHA survey that feature in the construction of our index of poor mental health and the questions from canonical depression and generalized anxiety disorder screeners (the PHQ-9 and General Anxiety Disorder-7 (GAD-7), respectively) known to be highly predictive of medical diagnoses (Kroenke, Spitzer, and Williams 2001; Spitzer et al. 2006). Online Appendix Figures A.14 and A.15 show that our index of poor mental health is strongly correlated with the PHQ-9 and GAD-7 scores (correlation coefficients of 0.66 and 0.61, respectively). The validation exercise is described in detail in online Appendix C.

Construction of the Treatment Indicator.—The construction of our treatment indicator is straightforward but for a minor caveat. A respondent to the NCHA survey is considered treated if, at the time the respondent took the survey, Facebook was available at her college and not treated otherwise. The caveat relates to the fact that we cannot determine whether or not a respondent was treated when the semester in which she took the survey coincides with the semester in which Facebook was rolled out at her college. For most of the analysis, we disregard such observations. In online Appendix A, we show that the results do not substantially change depending on whether we consider those respondents treated, untreated, or whether we assign them a treatment status of 0.5.

Identification Strategy.—The primary goal of this paper is to identify the causal impact of social media on mental health. A naïve correlation may be plagued by severe endogeneity concerns and, therefore, cannot credibly be given a causal interpretation. Examples of such endogeneity concerns include reverse causality (e.g., depressed individuals could use social media more) and omitted variable bias (e.g., the end of a romantic relationship might lead to both worse mental health outcomes and more free time to spend on social media).

To obtain estimates that can be more credibly interpreted as causal, we leverage the sharp and staggered rollout of Facebook across US colleges in the years 2004 through 2006. Under a set of assumptions described below, the quasi-experimental variation generated by the staggered Facebook rollout allows us to estimate the causal impact of social media on mental health using a generalized difference-in-differences strategy. The strategy compares the before-after difference in outcomes between students in colleges where Facebook was introduced and students in colleges that did not change their Facebook status between the two periods.

As a baseline specification, we estimate the following two-way fixed-effect (TWFE) model:

$$(1) \qquad Y_{icgt} = \alpha_g + \delta_t + \beta \times Facebook_{gt} + \mathbf{X}_i \times \gamma + \mathbf{X}_c \times \psi + \epsilon_{icgt},$$

where Y_{icgt} represents an outcome for individual i who participated in survey wave t and attends college c that belongs to expansion group g; α_g (or sometimes α_c) indicates expansion-group (or college) fixed effects; δ_t indicates survey-wave fixed

effects; $Facebook_{gt}$ is an indicator for whether, in survey wave t, Facebook was available at colleges in expansion group g; $\mathbf{X}_i$ and $\mathbf{X}_c$ are vectors of individual-level and college-level controls, respectively. We estimate equation (1) using ordinary least squares (OLS) and cluster standard errors at the college level.

To the extent that, in the absence of the Facebook rollout, the mental health outcomes of students attending colleges in different Facebook expansion groups would have evolved along parallel trends, and assuming college-level average treatment effects are homogeneous across treated colleges and over time, the coefficient of interest β identifies the average treatment effect on the treated (ATT) of the introduction of Facebook on student mental health.

Under the assumptions from the previous paragraph, the TWFE model allows us to rule out various concerns that could otherwise impair our ability to interpret the results as causal. First, we can rule out that the results are driven by time-invariant differences in mental health across colleges. Specifically, one could worry that more selective colleges recruit wealthier students who may have better (or worse) baseline mental health outcomes. By including Facebook-expansion-group or, depending on the specification, college fixed effects we can rule out such concerns. Second, we can rule out that our results are driven by mental health outcomes evolving over time in a way that is common across students at different colleges. For instance, certain macroeconomic fluctuations might influence all students' job prospects in a similar way, and, in turn, affect their mental health. Survey-wave fixed effects allow us to rule out such concerns.

One may worry about the plausibility of the parallel trends assumption in our setting. That is, one might worry that colleges belonging to different Facebook expansion groups might be on different mental health trends. We address this concern in four ways. First, we estimate a fully dynamic version of equation (1) and check for potential pretrends. Second, we explore the existence of pretrends by estimating a fully dynamic version of the alternative estimators introduced in De Chaisemartin and d'Haultfoeuille (2020); Borusyak, Jaravel, and Spiess (2021); Callaway and Sant'Anna (2021); and Sun and Abraham (2021). Third, to the extent that the trends are linear, we would be able to account for them in a robustness check that includes expansion-group-level linear time trends. Fourth, we present results using a specification that does not rely on our baseline college-level parallel trends assumption. In particular, we present results using a specification that includes college $\times$ survey-wave fixed effects and that compares students within the same college-survey wave who were exposed to Facebook for different lengths of time based on the year in which they entered college. These strategies, explored in detail in later sections, should assuage concerns about violations of the parallel trends assumption in our setting.

Limitations of TWFE Models and Suggested Remedies.—Although TWFE regressions similar to equation (1) are the workhorse models for staggered adoption research designs, they have been shown to deliver consistent estimates only under relatively strong assumptions about homogeneity in treatment effects (De Chaisemartin and d'Haultfoeuille 2020; Borusyak, Jaravel, and Spiess 2021; Callaway and Sant'Anna 2021; Goodman-Bacon 2021; Sun and Abraham 2021). Specifically, as shown in Goodman-Bacon (2021), the treatment effect estimate

obtained from a TWFE model is a weighted average of all possible 2×2 difference-in-differences comparisons between groups of units treated at different points in time. If treatment effects are homogeneous across treated groups and across time, the TWFE estimator is consistent for the ATT. Conversely, if treatment effects are heterogeneous across groups or time, the TWFE estimator does not deliver consistent estimates for the ATT.

We address concerns about the reliability of TWFE estimator by replicating our results using the robust estimators introduced in De Chaisemartin and d'Haultfoeuille (2020); Borusyak, Jaravel, and Spiess (2021); Callaway and Sant'Anna (2021); and Sun and Abraham (2021). By shutting down the 2×2 difference-in-differences comparisons between newly treated and already treated units, the robust estimators deliver consistent estimates even in the presence of heterogeneous treatment effects across time and/or treated units.

IV. Results

A. *Baseline Results*

Baseline Estimates.—Table 1 presents estimates of β in equation (1) on our overall index of poor mental health and shows that the introduction of Facebook at a college had a negative impact on student mental health. The first column in the table shows results for our simplest specification, which includes only Facebook-expansion-group fixed effects, survey-wave fixed effects, and an indicator for post-Facebook introduction. In the second column, we also include individual- and college-level control variables. In the third column, we replace Facebook-expansion-group fixed effects with college fixed effects to account for the changing composition of our sample. In the fourth column, we add expansion-group-level linear time trends, in order to take into account the possibility that colleges belonging to different Facebook expansion groups might be on different linear mental-health trends. Our results are fairly stable across specifications. The point estimates decrease but remain significant at the 5 percent level when college fixed effects and Facebook-expansion-group-level linear time trends are included.

The effect size on the index of poor mental health in our preferred specification, namely the one that includes college rather than Facebook-expansion-group fixed effects and that does not include linear time trends, is 0.085 standard deviation units. The effect above is estimated on the entire population of students taking the NCHA survey, which includes both students who did and who did not sign up for a Facebook account after Facebook was made available at their college. Therefore, the point estimate captures both the direct effect of Facebook on students who joined the platform and the indirect effect of Facebook on students who did not join the platform, but whose peers did. Although we cannot separate these two channels in the absence of data on an individual's Facebook use, we note that it is unlikely that our results are primarily driven by students who did not have a Facebook account.[19]

[19] As discussed in Section I, the average penetration rate of Facebook at each college was around 85 percent. Therefore, an effect concentrated solely among students who did not join the platform would have to be implausibly large (approximately 0.57 standard deviations in our main specification) to be consistent with our baseline effect.

Table 1—Baseline Results: Index of Poor Mental Health

	Index of poor mental health			
	(1)	(2)	(3)	(4)
Post-Facebook introduction	0.137 (0.040)	0.124 (0.022)	0.085 (0.033)	0.077 (0.032)
Observations	374,805	359,827	359,827	359,827
Survey-wave fixed effects	✓	✓	✓	✓
Facebook-expansion-group fixed effects	✓	✓		
Controls		✓	✓	✓
College fixed effects			✓	✓
FB-expansion-group linear time trends				✓

Notes: This table explores the effect of the introduction of Facebook at a college on student mental health. Specifically, it presents estimates of coefficient β from equation (1) with our index of poor mental health as the outcome variable. The index is standardized so that, in the preperiod, it has a mean of zero and a standard deviation of one. Column 1 estimates equation (1) without including controls; column 2 estimates equation (1) including controls; column 3, our preferred specification, replaces Facebook-expansion-group fixed effects with college fixed effects; column 4 includes linear time trends estimated at the Facebook-expansion-group level. Our controls consist of age, age squared, gender, indicators for year in school (freshman, sophomore, junior, senior), indicators for race (White, Black, Hispanic, Asian, Indian, and other), and an indicator for international student. Column 2 also includes indicators for geographic region of college (Northeast, Midwest, West, South); such indicators are omitted in columns 3 and 4 because they are collinear with the college fixed effects. For a detailed description of the outcome, treatment, and control variables, see online Appendix Table A.31. Standard errors in parentheses are clustered at the college level.

In order to help build intuition about the magnitude of our baseline effects, we provide a few benchmarks. First, the magnitude of our baseline effect corresponds to approximately 84 percent of the difference in the index of poor mental health between students in our sample with and without credit card debt. Second, we benchmark the magnitude of our estimates against the effect of a sudden unemployment spell on mental health. Comparing our estimates to the most closely related ones in a meta-analysis by Paul and Moser (2009), we find that the impact of introducing Facebook at a college on mental health is around 22 percent of the effect of job loss.[20] Third, we benchmark our results against the canonical PHQ-9 and GAD-7 mental health scales. We use data from the validation survey mentioned in Section III and discussed in detail in online Appendix C to determine how to weigh the variables contained in our index of poor mental health in a way that best predicts an indicator for having depression according to the PHQ-9 and an indicator for having generalized anxiety disorder according to the GAD-7. Next, we apply these weights to the NCHA sample to predict whether a student taking the NCHA survey would be classified as having depression or generalized anxiety disorder according to the PHQ-9 and GAD-7. Online Appendix Table A.30 shows that the introduction of Facebook increased by 2 percentage points the fraction of students

[20] Paul and Moser (2009) analyze studies estimating various aspects of mental health including symptoms of distress, depression, anxiety, psychosomatic symptoms, subjective well-being, and self-esteem. The estimates from Paul and Moser (2009) that can most credibly be interpreted as causal and hence be compared to our estimates are those that rely on quasi-experimental variation in job loss due to factory closures and mass layoffs.

whom, according to our prediction, the PHQ-9 and GAD-7 would classify as having depression or generalized anxiety disorder. The 2 percentage point increase corresponds to a 9 percent increase over the preperiod mean of 25 percent for depression and a 12 percent increase over the preperiod mean of 16 percent for generalized anxiety disorder.[21]

As a final benchmark, we leverage additional assumptions to compare our results to long-run mental health trends. The effect we find on the share of students who suffered from severe depression at least once in the last year is approximately 24 percent of the increase in that share between 2000 and 2019.[22] This number can be interpreted as the fraction of the increase in the prevalence of severe depression among college students that is explained by Facebook. Such calculation relies on strong assumptions and should therefore be interpreted with caution. Specifically, we assumed that (i) Facebook utilization rates among college students did not change substantially after 2004–2005; (ii) the effects of Facebook did not change over time; (iii) Facebook does not have cumulative effects.[23]

Figure 1 presents results on our individual outcome variables and shows that most of the dimensions of mental health in our dataset were negatively affected by the introduction of Facebook.[24] For all but one of the mental health outcomes from Figure 1, the point estimates are positive, which indicates worsened mental health. The conditions that appear to be most affected are depression and anxiety-related disorders, while the point estimates on anorexia and bulimia are close to zero.[25] The effect on severe depression is similar in magnitude to the effect observed in Allcott et al. (2020) on whether a respondent felt depressed in the past month (0.07 versus 0.09 standard deviations, respectively). This striking similarity is consistent with the possibility that the effects of the introduction of Facebook on depression are due primarily to direct use rather than general equilibrium effects. Having said that, the substantive differences between the studies, including the time period, target population, and empirical strategy, call for caution when drawing conclusions from such comparison.

The bottom section of Figure 1 also presents suggestive evidence that the introduction of Facebook at a college might have increased the extent to which students

[21] This exercise is discussed in more detail in online Appendix C.

[22] Data on the prevalence of severe depression among students come from ACHA reports containing aggregate statistics about mental health (ACHA 2000–2019). Since the wording of the question inquiring about severe depression changed in 2008 and caused a clear series break, we calculate the trend in depression by regressing the share of severely depressed students on year dummies, on whether the survey was conducted in the spring or fall, and on whether the survey contained the new wording. We define the trend in depression as the point estimate of the 2019 fixed effect dummy. According to our calculation, the share of severely depressed students increased by approximately 12 percentage points between 2000 and 2019. Based on our main specification, the introduction of Facebook at a college increased the share of students who reported suffering from severe depression at least once in the past year by 2.96 percentage points (p-value < 0.05). Hence, the effect of the introduction of Facebook is approximately 24 percent (2.96/12.15) of the increase in depression rates between 2000 and 2019.

[23] Section IVC, which shows that the negative effects of Facebook on mental health become stronger with longer exposure to the platform, already casts some doubt on assumption (iii).

[24] Online Appendix Table A.4 provides regression results for the individual mental health variables in both normalized (standard deviation) units and unnormalized (original) units. The table also provides unadjusted p-values and "sharpened" false discovery rate-adjusted q-values following the procedure of Benjamini, Krieger, and Yekutieli (2006), as outlined by Anderson (2008). The p-values are appropriate for readers with a priori interest in a particular outcome; the q-values adjust the inference for multiple hypotheses testing.

[25] Similar patterns can be observed in online Appendix Figure A.5 which is a version of Figure 1 with expansion-group-specific linear trends.

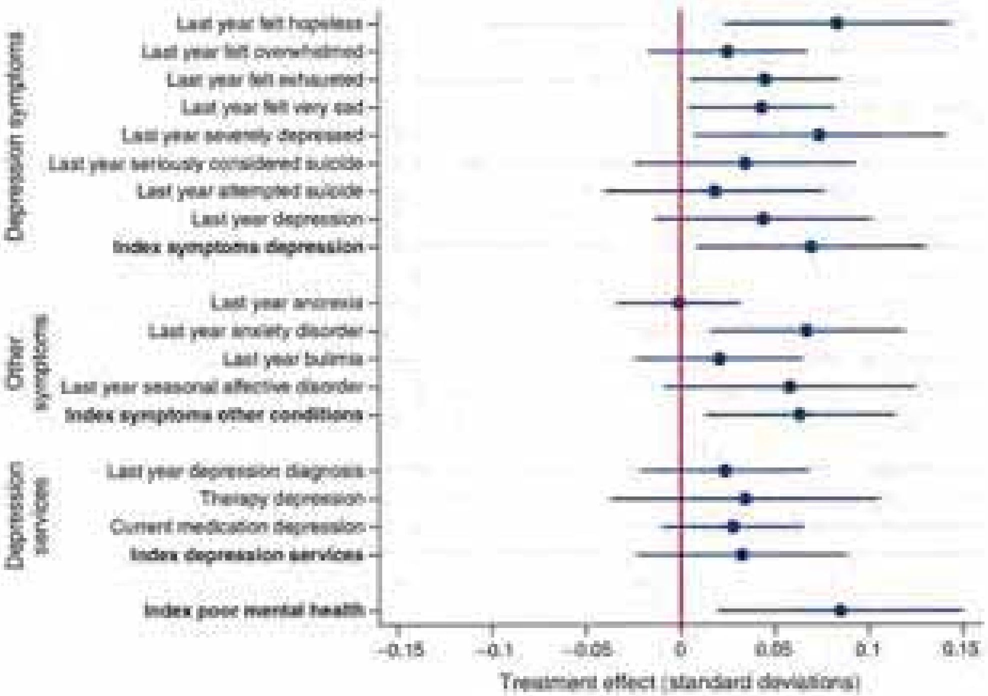


Figure 1. Effects of the Introduction of Facebook on Student Mental Health

Notes: This figure explores the effects of the introduction of Facebook at a college on all our mental-health outcome variables and on the related indices. Specifically, it presents estimates of coefficient β from equation (1) using our preferred specification, namely the one including survey-wave fixed effects, college fixed effects, and controls. The outcome variables are our overall index of poor mental health, the individual components of the index, and three subindices: the index of depression symptoms, the index of symptoms of other mental health conditions, and the index of depression services. All outcomes are standardized so that, in the preperiod, they have a mean of zero and a standard deviation of one. Our controls consist of age, age squared, gender, indicators for year in school (freshman, sophomore, junior, senior), indicators for race (White, Black, Hispanic, Asian, Indian, and other), and an indicator for international student. The reason why the point estimate on an index might be relatively large compared to the point estimates on each of the components of the index is that averaging across the index components reduces noise and, as a consequence, might increase the effect size measured in standard deviation units. For a detailed description of the outcome, treatment, and control variables, see online Appendix Table A.31. The bars represent 95 percent confidence intervals. Standard errors are clustered at the college level.

took up depression-related services. For all three items comprising the index of depression services (receiving an official depression diagnosis, going to therapy for depression, and taking antidepressants) the point estimates are positive, though not significant at conventional levels.[26] Finding a more muted average effect on depression-related services than on depression symptoms is arguably in line with intuition, in that an increase in symptoms of poor mental health induces the marginal student, rather than the average student, to take up mental healthcare services.[27] In Section IVB below, we show that students who, based on immutable baseline

[26] Note that, given the low average take-up of these services, the estimates represent large increases over the baseline mean. For antidepressants and psychotherapy, the point estimates represent an increase of about 13 percent and 20 percent over the baseline mean, respectively.

[27] The argument above relies on the baseline propensity to experience mental illness likely being normally distributed in the population (Plomin, Haworth, and Davis 2009) and the intuition that only individuals above a

characteristics, are predicted to be most susceptible to mental illness—and therefore more likely to be on the margin of receiving a depression diagnosis—are indeed significantly more likely to take up depression-related services after the introduction of Facebook.

Event Study Figures.—In order to test for parallel trends and study the dynamics of treatment effects, we estimate an event-study version of the TWFE model with indicators for distance to/from the introduction of Facebook. Specifically, we estimate the following specification:

$$Y_{igt} = \alpha_g + \delta_t + \beta_k \times \sum_{k=-8}^{5} D_{k(gt)} + \epsilon_{igt}, \tag{2}$$

where Y_{igt} is our index of poor mental health and $D_{k(gt)}$ is set of indicator variables that take value one if, for expansion group g in survey wave t, the introduction of Facebook was k semesters away. When estimating the model using OLS, we treat students who took the survey in the semester just before Facebook was rolled out at their college as the omitted category and compare them to students who took the NCHA survey in other semesters.

As discussed in Sun and Abraham (2021), the fully dynamic version of the TWFE model in equation (2) estimated using OLS delivers consistent estimates only under relatively strong assumptions regarding treatment effect homogeneity. In order to allow for heterogeneity in treatment effects across time and treated units, we also present the event study figures generated by a set of recently proposed estimators that are robust to treatment effect heterogeneity (De Chaisemartin and d'Haultfoeuille 2020; Borusyak, Jaravel, and Spiess 2021; Callaway and Sant'Anna 2021; Sun and Abraham 2021).

Figure 2 presents the event-study figures and shows that the estimates are consistent with the parallel trends assumption: independently of the estimator used, the coefficients on the semesters prior to the introduction of Facebook at a college are all close to zero and exhibit no discernible pretrends.[28] Figure 2 also sheds light on the dynamics of treatment effects: all the recently developed robust estimators show treatment effects that increase over time in the postperiods.[29] The increase in treatment effects over time could be explained by (i) higher adoption rates at a college over time; (ii) higher intensity of usage at the individual level over time; (iii) the effects becoming stronger as a function of length of exposure to the platform. Given the evidence presented in Section I on the rapid and widespread penetration of Facebook at each college and evidence that intensity of usage did not

certain threshold in the right tail of the distribution experience sufficiently severe symptoms to seek out mental healthcare services.

[28]Online Appendix Figure A.4 shows the TWFE OLS estimates of a version of equation (2) that considers each of the first three Facebook expansion groups in turn and compares it to the last Facebook expansion group. These figures are constructed at the yearly level to reduce noise arising from the smaller number of observations. Consistent with Facebook having a negative impact on student mental health, in all the pairwise comparisons, all the estimates in the postperiod are positive and most are statistically significant while the estimates in the preperiod are not statistically different from zero.

[29]Contrary to the recently developed robust estimators, the OLS estimator shows a relatively flat trend in the postperiod. This is likely because, in the case of dynamically increasing treatment effects, the OLS estimator, which uses already treated units as controls for newly treated units, exhibits a downward bias.

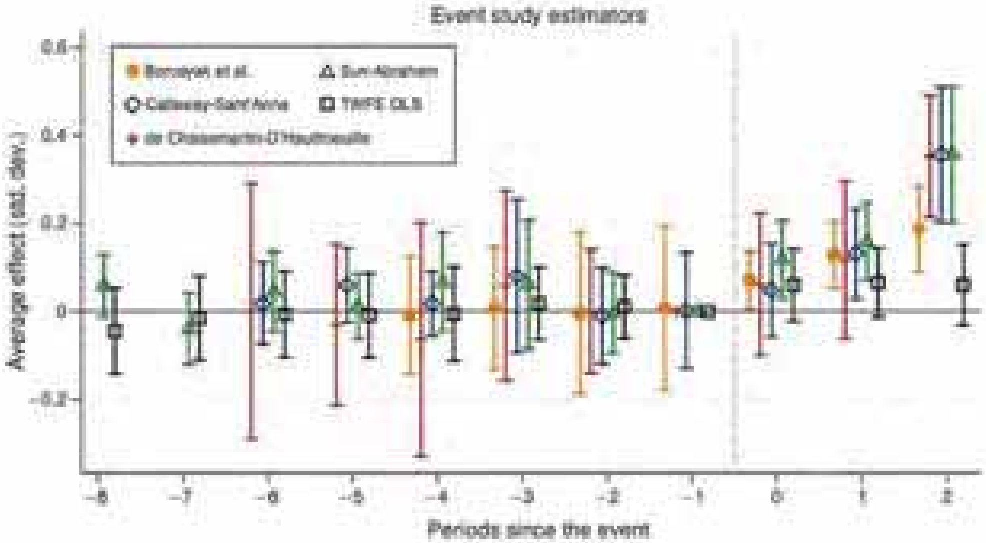


Figure 2. Effects of Facebook on the Index of Poor Mental Health Based on Distance to/from Facebook Introduction

Notes: This figure overlays the event-study plots constructed using five different estimators: a dynamic version of the TWFE model, equation (2), estimated using OLS (in black with square markers); Sun and Abraham (2021) (in green with triangle markers); Callaway and Sant'Anna (2021) (in blue with diamond markers); De Chaisemartin and d'Haultfoeuille (2020) (in red with cross markers); and Borusyak, Jaravel, and Spiess (2021) (in orange with circle markers). The outcome variable is our overall index of poor mental health. The time variable is the survey wave and the treatment group variable is given by the semester in which the college attended by the student was granted Facebook access. The figure displays only two postperiods because the estimation of additional post periods would require employing already treated units as controls for newly treated units. In the presence of heterogeneous dynamic treatment effects, such comparisons would bias the estimation and, therefore, they are shut down by all the newly introduced robust estimators. As a result, the maximum number of postperiods that can be estimated robustly is two. For the Borusyak, Jaravel, and Spiess (2021) estimator, we estimate four preperiods since estimating more preperiods dramatically increases the standard errors in the preperiod (Borusyak, Jaravel, and Spiess 2021, p. 24). Similarly, for the estimator by De Chaisemartin and d'Haultfoeuille (2020), the maximum number of preperiods that can be estimated in our panel is only five. In order to estimate the standard errors for the $t+2$ estimate, the De Chaisemartin and d'Haultfoeuille (2020) estimator includes controls for age and age squared. For appropriate estimation of the coefficients on $t=-8$ and $t=-7$ using the Sun and Abraham (2021) estimator, we include data from additional preperiods, even though, in those preperiods, we do not observe all four Facebook expansion groups (Sun and Abraham 2021, p. 13). For a detailed description of the outcome and treatment variables, see online Appendix Table A.31. The bars represent 95 percent confidence intervals. Standard errors are clustered at the college level.

increase substantially over time (Stutzman 2006; Lampe, Ellison, and Steinfield, 2008), we tentatively lean in favor of the length-of-exposure explanation. We further study the effects of differential length of exposure to Facebook at the individual level in Section IVC.

B. *Heterogeneity*

Heterogeneity by Predicted Susceptibility to Mental Illness.—In order to study whether the introduction of Facebook at a college led students on the margin of a depression diagnosis to take up depression-related services, we proceed in two steps: first, we estimate a least absolute shrinkage and selection operator (LASSO) to identify individuals who, based on baseline immutable characteristics, are more

susceptible to mental illness. Second, we show heterogeneous treatment effects based on our LASSO-predicted measure of susceptibility to mental illness.

The LASSO prediction is generated as follows: first, we construct an indicator that equals one if a student has ever been diagnosed with a mental health condition. Second, we consider a set of immutable individual-level characteristics (age, year in school, gender, race, an indicator for US citizenship, and height), generate all two-way interactions between these characteristics, and generate second- and third-order monomials of each characteristic. Third, we implement a LASSO procedure in the preperiod to predict our indicator for ever having been diagnosed with a mental health condition based on the immutable individual-level characteristics and functions thereof described above.

In order to test the quality of the prediction, we plot our measure of predicted susceptibility to mental illness against our index of poor mental health. Online Appendix Figure A.7 shows a strong relationship between the index of poor mental health and our predicted measure of susceptibility to mental illness.

Armed with our LASSO prediction, we can study how the introduction of Facebook at a college affected students across the mental-illness-susceptibility spectrum, and whether it induced students who are more likely to be on the margin of a depression diagnosis to seek out depression-related services such as psychotherapy. The upper left panel of Figure 3 presents the estimated effects on the index of poor mental health across quintiles of our LASSO-predicted measure of susceptibility to mental illness.[30] As shown in the figure, the effects of the introduction of Facebook on symptoms of poor mental health tend to be stronger for individuals with a higher baseline risk of developing mental illness.[31]

The upper right panel of Figure 3 shows that the introduction of Facebook on the take-up of depression-related services exhibits a similar pattern. We find weak positive effects throughout the distribution of predicted susceptibility to mental illness, though for most quintiles the point estimates are fairly small and not statistically significant. The effects become more pronounced for individuals in the top quintile; in particular, the point estimate on the top quintile is relatively large in magnitude (0.063 standard deviations) and four times as large as the point estimate on the bottom quintile. As indicated in column 2 of online Appendix Table A.5, the difference between the coefficients for the top and the bottom quintiles is significant at the 1 percent level. These results suggest that, indeed, students who are predicted to be most susceptible to mental illness—and therefore more likely to seek mental healthcare

[30] Specifically, we estimate the following modification of equation (1):

$$(3) \quad Y_{icgt} = \alpha_c + \delta_t + \beta_q \times Facebook_{gt} \times MHSusceptQ_i + \zeta \times MHSusceptQ_i + \mathbf{X}_i \times \gamma + \mathbf{X}_c \times \psi + \epsilon_{icgt},$$

where $MHSusceptQ_i$ are the quintiles of i's predicted susceptibility to mental illness. Figure 3 presents the estimates of β_q. Online Appendix Table A.5 presents these estimates in a table form, together with p-values for comparisons between the first quintile and other quintiles.

[31] We note that, for predicting baseline susceptibility to mental illness, the stock variable of "having ever been diagnosed" with a mental illness is arguably more relevant than the flow variable of having exhibited a certain symptom in the past year, because the former captures information covering a longer time span. Online Appendix Figure A.10 examines robustness of our results to an alternative measure of susceptibility to mental illness based on a LASSO regression predicting whether a respondent's index of poor mental health is in the top 10 percent of the preperiod sample. The results are qualitatively similar. As shown in a corresponding online Appendix Table A.6, the coefficient for the top quintile remains statistically different from the coefficient for the bottom quintile at the 10 percent level for all outcomes.

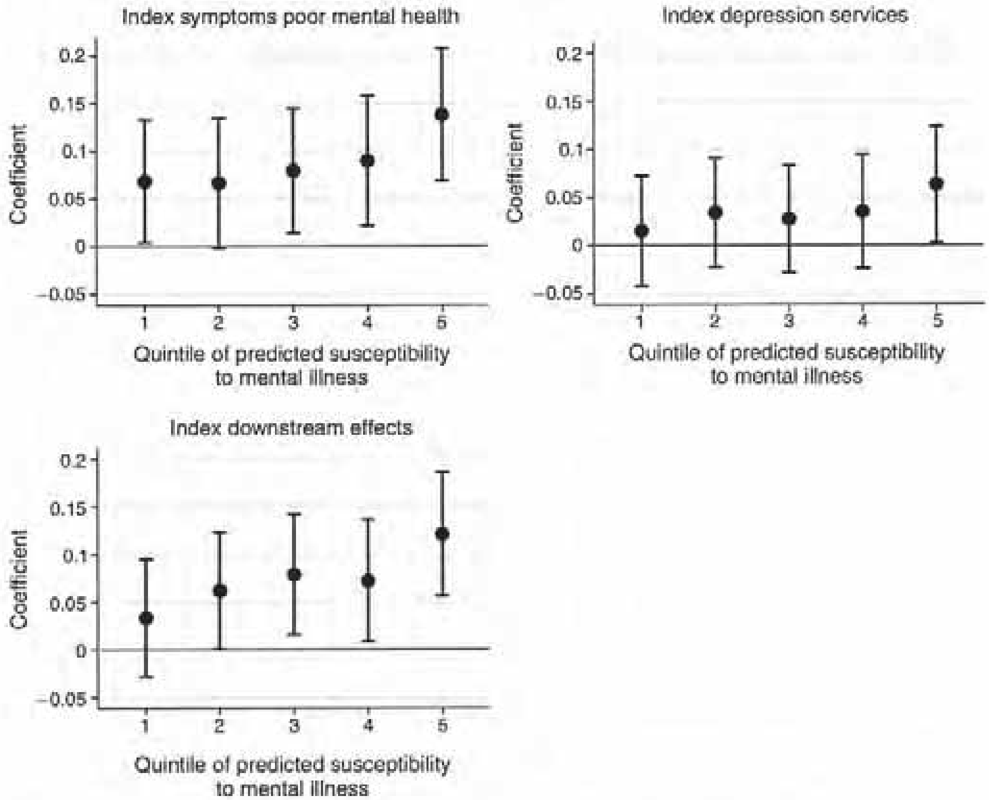


FIGURE 3. HETEROGENEOUS EFFECTS BY PREDICTED SUSCEPTIBILITY TO MENTAL ILLNESS

Notes: This figure explores the extent to which the effects of the introduction of Facebook at a college are heterogeneous depending on students' predicted susceptibility to mental illness. Specifically, it presents the estimates from equation (3) in which our indicator for post-Facebook introduction is interacted with a set of indicators for belonging to each quintile of a LASSO-predicted measure of susceptibility to mental illness. The outcome variable in the top-left panel is our index of symptoms of poor mental health; the outcome variable in the top-right panel is our index of depression services; the outcome variable in the bottom-left panel is our index measuring whether students reported that conditions related to poor mental health negatively affected their academic performance. All indices are standardized so that, in the preperiod, they have a mean of zero and a standard deviation of one. The estimates (also displayed in online Appendix Table A.5) are obtained using our preferred specification, namely the one including survey-wave fixed effects, college fixed effects, and controls. Our controls consist of age, age squared, gender, indicators for year in school (freshman, sophomore, junior, senior), indicators for race (White, Black, Hispanic, Asian, Indian, and other), and an indicator for international student. For a detailed description of the outcome, treatment, interaction, and control variables, see online Appendix Table A.31. The bars represent 95 percent confidence intervals. Standard errors are clustered at the college level.

due to a worsening in symptoms—are more likely to take up depression-related services such as psychotherapy for depression and antidepressants as a result of the introduction of Facebook.

Other Dimensions of Heterogeneity.—Online Appendix Figure A.6 estimates heterogeneous effects across several baseline characteristics. Consistent with surveys showing that women use social media more often and are more likely to report using Facebook for longer than they intend, we find suggestive evidence that the

results are larger among women (Thompson and Lougheed 2012; Lin et al. 2016).[32] We also find stronger effects on non-Hispanic Whites, and a weaker effect on international students, younger students, and first-years.

C. *Effects Based on Length of Exposure to Facebook*

The effects of the introduction of Facebook estimated thus far leverage variation that occurs at the college-survey-wave level. Our dataset also features variation at the college-survey-wave–year-in-school level that we can leverage to study the effects of length of exposure to Facebook at the level of individual students. For instance, in the early spring of 2006, a freshman at Harvard would have been exposed to Facebook for one full semester, whereas a senior at Harvard would have been exposed for more than three full semesters.

In order to study the effects of length of exposure to Facebook at the level of individual students, we estimate a version of equation (1) with individual-level treatment intensity. In this alternative specification, we include a student-level treatment component that equals the number of semesters that the student had access to Facebook given: (i) the college the student attends; (ii) the survey wave the student participated in; and (iii) the year in which the student started college. Specifically, we estimate the following equation:

$$Y_{icgt} = \alpha_c + \delta_t + \sum_{k=0}^{5} \beta_k \times Semesters_{k(ict)} + \mathbf{X}_i \times \gamma + \epsilon_{icgt}, \tag{4}$$

where $Semesters_{k(ict)}$ is a set of indicators that equal one if student i at college c in survey-wave t had access to Facebook for k semesters. The number of treated semesters is calculated as $k = FB_{gt} \times (t - \max\{\tau_i, \tau_c\})$; t represents time in semesters; τ_c represents the semester in which Facebook was introduced at college c attended by student i; τ_i represents the semester in which student i started studying at college c; and, as before, FB_{gt} is the indicator function for whether Facebook was available at student i's college c by time t.[33]

Figure 4 displays the estimates of β_k and shows that the negative effects of the introduction of Facebook on mental health worsen the longer students are exposed to Facebook. Online Appendix Table A.7 presents the results in a regression framework where we assume that the effects grow linearly over time. The table shows that the number of treated semesters has a significant effect on our main index, on symptoms of poor mental health, and on the utilization of depression-related healthcare services.

Since the index of depression services only comprises binary variables that have a straightforward yes/no interpretation, we provide intuition for the magnitude of our results by presenting the effects on each component of the index of poor

[32] Furthermore, baseline prevalence of depression is found to be higher among women, across different nations, cultures, and age groups (Nolen-Hoeksema and Hilt 2008; Albert 2015; Salk, Hyde, and Abramson 2017). Thus, the slightly stronger effects among women are also consistent with studies showing that women are more likely to be affected by certain mental illnesses.

[33] Students who entered college in 2006 might have been exposed to Facebook already in high school, because, starting in September 2005, college students with Facebook access could invite high school students to join the platform. We exclude cohorts of students who might have been exposed to Facebook in high school from the length-of-exposure analysis. Including them does not meaningfully affect the results.

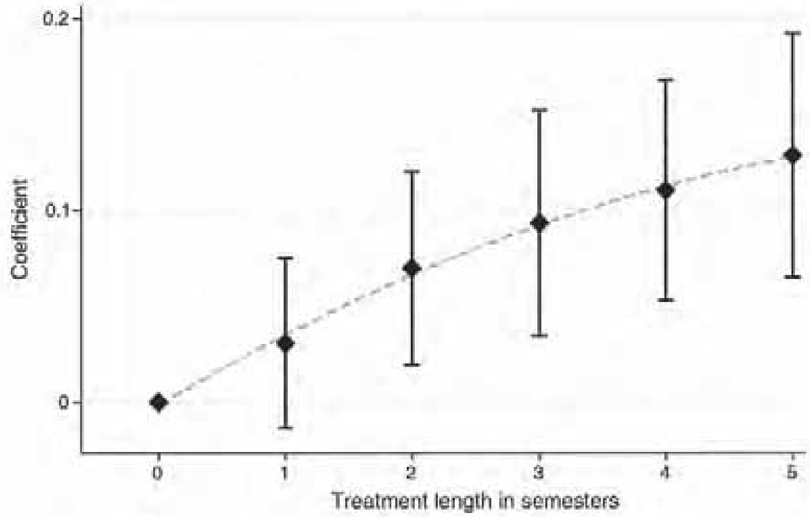


FIGURE 4. EFFECT ON POOR MENTAL HEALTH BY LENGTH OF EXPOSURE TO FACEBOOK

Notes: This figure explores the effects of length of exposure to Facebook on our index of poor mental health by presenting estimates of equation (4). The index is standardized so that, in the preperiod, it has a mean of zero and a standard deviation of one. The dashed curve is the quadratic curve of best fit. Our controls consist of age, age squared, gender, indicators for year in school (freshman, sophomore, junior, senior), indicators for race (White, Black, Hispanic, Asian, Indian, and other), and an indicator for international student. Students who entered college in 2006 might have been exposed to Facebook already in high school, because, starting in September 2005, college students with Facebook access could invite high school students to join the platform. Such students are excluded from the regression. For a detailed description of the outcome, treatment, and control variables, see online Appendix Table A.31. The bars represent 95 percent confidence intervals. Standard errors are clustered at the college level.

mental health services in original units. Specifically, online Appendix Table A.8 shows that being exposed to Facebook for five semesters increases the probability that a student is diagnosed with depression by around 32 percent, the probability that a student is in therapy for depression by around 50 percent, and the probability that a student is on antidepressants by around 33 percent.

D. *Robustness Checks and Alternative Explanations*

Robustness Checks.—Online Appendix A describes a battery of exercises that probe the robustness of our estimates. The exercises include various placebo tests on variables that should not be affected by the introduction of Facebook and modified versions of our main specifications that take into account a host of possible concerns related to (i) the construction of our index of poor mental health, (ii) the construction of our treatment variable, (iii) particular Facebook expansion groups driving the effects, (iv) particular variables driving the effects, (v) the parallel trends assumption, and (vi) the level at which standard errors are clustered. We highlight one of our most convincing robustness check, which consists of a variant of the length-of-exposure specification from Section IVC that includes college by survey-wave fixed effects. Such specification, which delivers estimates consistent with the hypothesis that longer exposure to Facebook has a negative impact on

student mental health, does not rely on our baseline college-level parallel trends assumption for identification.

Stigma as an Alternative Explanation.—One might worry that Facebook affected the stigma associated with mental illness and that our results may not reflect an increase in the prevalence of mental illness per se but rather an increase in willingness to discuss it. To formally investigate the role of stigma, we adopt a three-pronged strategy. First, we collected all the college newspaper articles containing the word Facebook published around the time of Facebook's expansion and checked whether any of them mention stigma in relation to mental health. While we do find articles hinting at potential negative effects of Facebook on mental health, we do not find any articles mentioning stigma. Second, we study whether the fraction of missing answers to the mental health questions in the NCHA survey was affected by the introduction of Facebook. If Facebook made people more comfortable discussing mental illness, we would expect to observe fewer missing answers after the introduction of Facebook.[34] Consistent with the effects being driven by increased prevalence of mental illness rather than by stigma, online Appendix Table A.18 shows that the prevalence of missing answers was not affected by the introduction of Facebook. Third, in Section V, we present evidence that the introduction of Facebook did not affect the reporting of other stigmatized conditions, such as being a victim of sexual assault or consuming illegal drugs. Furthermore, we find no detectable effects of the introduction of Facebook on eating disorders, even though such conditions are often highly stigmatized (Puhl and Suh 2015). If reduction in stigma was indeed the driving force behind our results, it would be surprising not to find similar effects on other stigmatized behaviors and conditions.

E. *Downstream Implications of Poor Mental Health*

Does the effect of Facebook on mental health have negative downstream repercussions on academic performance? According to the students' reports, the answer is affirmative.

One of the NCHA survey questions inquires as to whether various conditions affected the students' academic performance. The conditions related to mental health are attention deficit disorder, depression/anxiety disorder/seasonal affective disorder, eating disorders, stress, and sleep difficulties.[35] The main advantage of analyzing these questions is that they trace a pathway from the introduction of Facebook to perceptions of worsened academic performance via poor mental health. It is important to emphasize, however, that we do not directly measure effects on grades, and that we do not rule out potential positive effects of Facebook on students' academic performance due to channels unrelated to mental health, such as improved teamwork.[36]

[34] Indeed, missing values are more common in the NCHA survey among sensitive questions (Kays, Gathercoal, and Buhrow 2012).

[35] According to the DSM-5, sleep difficulties are a symptom of depression (APA 2013). Similarly, stress has been associated with depression (Yang et al. 2015).

[36] The NCHA dataset does include a question inquiring about the students' cumulative GPA, but the effects of the introduction of Facebook on cumulative GPA are small and noisy. This is likely because the answer options to

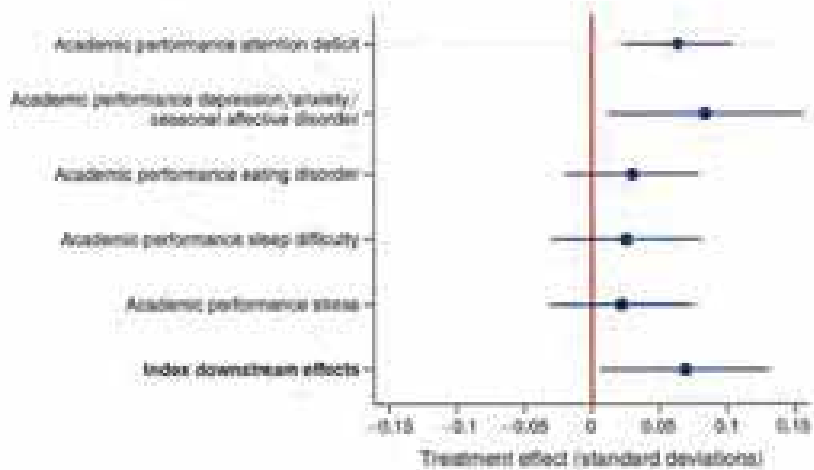


Figure 5. Downstream Effects on Academic Performance

Notes: This figure explores downstream effects of the introduction of Facebook on the students' academic performance. It presents estimates of coefficient β from equation (1) using our preferred specification, including survey-wave fixed effects, college fixed effects, and controls. The outcome variables are answers to questions inquiring as to whether various mental health conditions affected the students' academic performance and our index of downstream effects. All outcomes are standardized so that, in the preperiod, they have a mean of zero and a standard deviation of one. For a detailed description of the outcome, treatment, and control variables, see online Appendix Table A.31. The bars represent 95 percent confidence intervals. Standard errors are clustered at the college level.

Figure 5 presents estimates of equation (1) and shows how the introduction of Facebook affected each of the measures described in the previous paragraph. All the point estimates are positive and the coefficient for an equally weighted index summarizing them is positive and significant, suggesting that, after the introduction of Facebook, students were more likely to report that their academic performance was impaired as a result of poor mental health. The effect size on the index is 0.067 standard deviation units. Consistent with our evidence suggesting that depression and anxiety-related disorders are the conditions most severely affected by the introduction of Facebook, we find the largest effect on the depression/anxiety-disorder/seasonal-affective-disorder measure. The number of students who reported that those conditions impaired their academic performance increased by 3 percentage points over a baseline of 13 percent. Finally, the bottom-left panel in Figure 3 and column 3 of online Appendix Table A.5 show that the negative effect of poor mental health on self-reported academic performance is especially pronounced among the students who are predicted to be most susceptible to mental illnesses.

the GPA question are rather course (A, B, C, D/F), because cumulative GPA is a stock variable whose value might largely be determined before the introduction of Facebook at a college, and because students might receive grades based on relative rather than absolute performance. We note that, when analyzing questions on how mental health conditions affected academic performance, it is possible to find an effect even if students are graded on a curve. In particular, students' absolute performance and perception thereof can decrease as a result of the introduction of Facebook.

V. Mechanisms

Recent scholarship identified two main channels whereby Facebook might directly affect mental health: unfavorable social comparisons (Appel, Gerlach, and Crusius 2016) and disruptive internet use (Griffiths, Kuss, and Demetrovics 2014). Another, albeit indirect, possibility is that the introduction of Facebook might lead to behavioral changes that, in turn, affect mental health. We present evidence related to each set of mechanisms in turn. Overall, our evidence is mostly consistent with the unfavorable social comparisons channel.

Unfavorable Social Comparisons.—Facebook and other social media platforms make it easier for people to compare themselves to members of their social networks.[37] Such social comparisons, if unfavorable, could be detrimental to users' self-esteem and mental health (Vogel et al. 2014).[38]

Theoretically, the set of individuals who might be negatively affected by social comparisons is unclear. A simple model of social comparisons might posit that individuals compare themselves to the median member of their group along some dimension of interest (e.g., popularity, wealth, or looks).[39] If social media users are sophisticated, they will be able to extract accurate information from social media platforms about their relative ranking vis-à-vis their peers along the dimension of interest. In that case, we might expect around half of social media users to benefit from social comparisons and about half to suffer from them. Conversely, if social media users are to some extent naïve, they will fail to understand that the content that their peers post on social media is likely to be highly curated rather than representative (Appel, Gerlach, and Crusius 2016). In that case, they will systematically underestimate their relative ranking vis-à-vis their peers and, as a result, more than half of them will perceive social comparisons on Facebook as unfavorable.

In this section, we present evidence showing that (i) subpopulations which, in virtue of their baseline characteristics, might be more likely to suffer from social comparisons exhibit larger effects;[40] (ii) the introduction of Facebook did not correct the students' misperceptions about their peers' social lives and, in some cases, exacerbated them. The latter piece of evidence is consistent with students exhibiting a degree of naïveté in interpreting the information conveyed through social media.

Figure 6 shows that the introduction of Facebook at a college affected more severely the mental health of students who might be more likely to be affected by unfavorable social comparisons. The figure plots estimates of the coefficient on the interaction between our treatment indicator and various moderators in a regression

[37] Indeed, surveys reveal that college students generally used Facebook to learn more about their classmates or about individuals they already knew offline, and used it less often to meet new people (Lampe, Ellison, and Steinfield 2008).

[38] We consider "fear of missing out" (FoMO) as being related to social comparisons, though we recognize that certain features of the phenomenon may not be fully captured by social comparisons. In relation to social media, FoMO refers to the idea that social media platforms might make users more aware of the existence of exciting events that they are missing out on.

[39] Individuals could compare themselves to some other percentile of the distribution. The higher the percentile, the larger the set of individuals who would suffer from an increase in the ability to engage in social comparisons.

[40] Such subpopulations are expected to exhibit larger effects independently of whether, in general, social media users are naive or sophisticated.

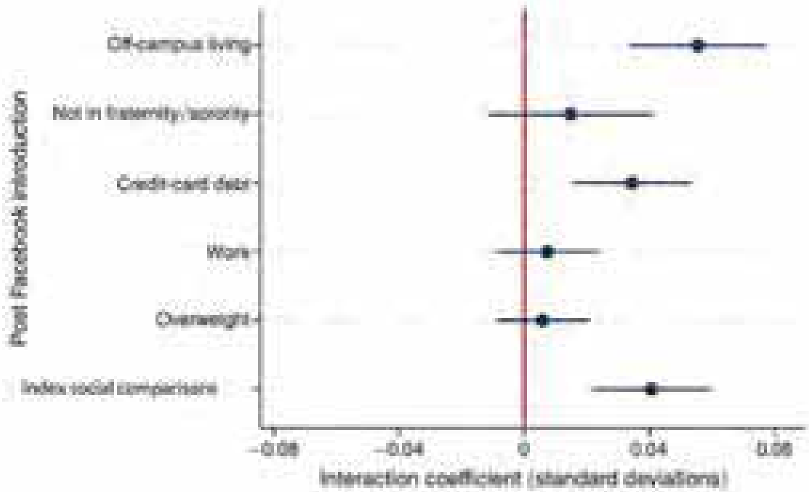


Figure 6. Heterogeneous Effects as Evidence of Unfavorable Social Comparisons

Notes: This figure explores the mechanisms behind the effects of Facebook on mental health. It presents estimates from a version of equation (1) in which our treatment indicator is interacted with a set of indicators for belonging to a certain subpopulation of students. The outcome variable is our overall index of poor mental health. The estimates are obtained using our preferred specification, namely the one including survey-wave fixed effects, college fixed effects, and controls. For a detailed description of the outcome, treatment, interaction, and control variables, see online Appendix Table A.31. The bars represent 95 percent confidence intervals. Standard errors are clustered at the college level.

with our index of poor mental health as the outcome variable. Specifically, we consider the following subpopulations of students: (i) students who live off campus and are therefore less likely to participate in on-campus social life, (ii) students who have weaker offline social networks as measured by not belonging to a fraternity or sorority organization, (iii) students who have lower socioeconomic status as measured by carrying credit card debt or working part-time alongside studying, and (iv) students who are overweight. We generate an index of social comparisons based on the variables above and consider, as an additional moderator, an indicator that takes value one if a student is above the median value of the index of social comparisons. All of the point estimates are positive and we find a strong and statistically significant effect on the index, on students living off campus, and on students with credit card debt. Consistent with the social comparison mechanism, the introduction of Facebook has particularly detrimental effects on the mental health of students who might view themselves as comparing unfavorably to their peers.[41]

To test whether the introduction of Facebook affected the students' beliefs about their peers' social lives, we estimate the impact of the rollout of Facebook on all survey questions that elicit students' perceptions of their peers' drinking

[41] Of course, we cannot rule out that the subpopulations above exhibit larger effects for reasons other than social comparisons. One concern we can rule out is that such subpopulations exhibit larger effects because they have worse baseline mental health. Online Appendix Figure A.11 shows a version of Figure 6 in which we include as an additional control our treatment indicator interacted with our individual-level LASSO-predicted measure of susceptibility to mental illness. The results are not meaningfully affected.

behaviors.[42] Specifically, we study the following three sets of beliefs: (i) beliefs about the number of alcoholic drinks the typical student has at a party, (ii) beliefs about the share of the student population who has had an alcoholic drink in the month before the survey, and (iii) beliefs about the share of the student population who drinks alcohol on a regular basis. Online Appendix Table A.19a finds a positive and significant effect on each of the three outcomes above and on an equally weighted index summarizing the three outcomes. Furthermore, online Appendix Table A.20 shows that the effects on perceptions are particularly pronounced for students who live off campus and who, therefore, have to rely more heavily on social media when estimating their peers' behaviors.[43]

Did Facebook affect beliefs about alcohol consumption because it led students to actually drink more, or did Facebook affect beliefs without a concurrent increase in drinking behaviors? Online Appendix Table A.19b shows that the effects on self-reported alcohol usage are substantially smaller than the effects on perceptions, suggesting that the effects on perceptions are unlikely to be driven by a change in actual behavior.[44]

If peers' behaviors did not change, why did Facebook affect perceptions? One option is that baseline perceptions were incorrect and the additional information provided on Facebook corrected such misperceptions. An alternative explanation is that Facebook led students to update their beliefs, but without aligning them more closely to reality. Online Appendix Table A.21 shows that the introduction of Facebook at a college did not lead students to develop more accurate perceptions about their peers' drinking behaviors and, for one of the outcomes, significantly exacerbated misperceptions. Specifically, the table estimates the effects on the difference between a student's perception of the alcohol consumption of the typical student at her college and the actual typical consumption at the student's college calculated using self-reported alcohol usage in the student's college-survey wave. The results are consistent with students failing to fully take into account the fact that the content they see on social media is a curated rather than representative portrayal of their peers' lives. Such naïveté could lead to distorted beliefs and exacerbate the effects of social comparisons.[45]

[42]We focus on drinking behavior because alcohol is the most commonly consumed intoxicant among college students and because the NCHA survey includes several questions on drinking-related perceptions.

[43]Online Appendix Table A.25 provides suggestive evidence that perceptions regarding other students' sexual behavior may have also been affected by the introduction of Facebook. Conversely, online Appendix Table A.27 shows that perceptions regarding the usage of illicit substances did not change. Finding effects on the perceptions of alcohol consumption but not on the perceptions of drug consumption is consistent with the fact that drinking and positive references to alcohol were common on Facebook profiles at the time, whereas images of students using drugs were very rare (Watson, Smith, and Driver 2006; Kolek and Saunders 2008; Morgan, Snelson, and Elison-Bowers 2010).

[44]If the introduction of Facebook decreased the stigma related to alcohol consumption, our results about alcohol usage could be biased (see also our discussion of stigma in the context of mental health in Section IVD). Although we cannot rule out the possibility that changes in stigma due to the introduction of Facebook had an effect specifically on alcohol-related questions, such bias would, if anything, make our results even starker. Specifically, if the introduction of Facebook reduced the stigma around underage drinking, the actual effect on alcohol usage would be smaller than the effect we estimate. Thus, the gap between the changes in usage and changes in perceptions would be even larger than the effect we currently estimate.

[45]Although it is easy to imagine that Facebook users might learn over time how to interpret the content they are exposed to on social media, a recent review of the psychology literature points to social comparisons as a concern that is relevant to this day (Verduyn et al. 2020).

Disruptive Internet Use.—The second direct channel whereby social media may negatively affect mental health is disruptive internet use (Griffiths, Kuss, and Demetrovics 2014). Specifically, some scholars argue that social media use might disrupt concentration, impair people's ability to focus, and lead to anxiety (e.g., Paul, Baker, and Cochran 2012; Meier, Reinecke, and Meltzer 2016).

We do not find significant evidence supporting the disruptive internet use channel. The main survey question that speaks to disruptive internet use asks students whether the internet or computer games affected their academic performance. Students could answer that the issue affected their academic performance, that they experienced the issue but it did not affect their performance, and that they did not experience the issue. If, after the introduction of Facebook at their college, students found the internet more distracting and had a harder time focusing because of it, we would expect a larger number of students to answer that they experienced the internet or computer games as an issue and that it affected their academic performance. Online Appendix Table A.22 shows that the share of students experiencing internet or computer games as an issue increased by around 5 percent, but the effect is not statistically significant.

Other Behaviors.—The introduction of Facebook at a college might have led students to engage or refrain from engaging in a set of other behaviors that have some bearing on mental health. For instance, the rollout of Facebook might have popularized illicit drug use.

Online Appendix Tables A.23, A.24, and A.26 present estimates of the effects of the introduction of Facebook using equation (1) on various offline behaviors measured in the survey that could plausibly affect mental health. Comfortingly, we do not find any effects on sexual assaults. Similarly, none of the outcomes related to relationships and drug use exhibit significant effects. Combined with the null results on drinking behaviors (online Appendix Table A.19b), we do not find much evidence that the introduction of Facebook at a college had meaningful effects on various self-reported behaviors that could have a bearing on mental health.

VI. Discussion

In this section, we elaborate on the extent to which our findings have the potential to inform our understanding of the effects of social media on mental health today.

Over the last two decades, Facebook underwent a host of important changes. Such changes include (i) the introduction of a personalized feed where posts are ranked by an algorithm, (ii) the growth of Facebook's user base from US college students to almost three billion active users around the globe (Facebook 2021), (iii) video often replacing images and text, (iv) increased usage of Facebook on mobile phones instead of computers, and (v) the introduction of Facebook pages for brands, businesses, and organizations. The nature of the variation we are exploiting does not allow us to identify the impact of these features of social media. For instance, our estimates cannot shed light on whether the increased reliance on Facebook for news consumption has exacerbated or mitigated the effects of Facebook on mental health.

Similarly, we cannot provide evidence as to whether years of experience with the platform mitigate or exacerbate the effects on mental health.[46]

Despite these caveats, we believe the estimates presented in this paper are still highly relevant today for two main reasons. First, the mechanisms whereby social media use might affect mental health arguably relate to core features of social media platforms that have been present since inception and that remain integral parts of those platforms today. At their core, Facebook and similar platforms are online forums where individuals share information, often about themselves, including pictures, videos, and personal details. Even today, the most common primary reason for using social media is staying in touch with family and friends, in contrast to reading news stories or watching live streams (GWI 2021). The ease with which one can access information about ones' network, together with the fact that the content posted on social media is generally highly curated, might naturally invite social comparisons. To the extent that the effects of Facebook on mental health at inception were at least partly driven by unfavorable social comparisons, we would expect our findings to still be relevant today.

Second, the mechanisms whereby Facebook use can affect mental health might have been exacerbated rather than mitigated by many of the technological changes undergone by Facebook and related platforms in the last 15 years. Individuals now receive information about their social network directly in their news feeds, and the information is more relevant to them because it is ranked by an algorithm. The content on the platform is richer in that it often includes videos, and it can be accessed at any time or place using a smartphone. These changes might make Facebook even more engaging and might exacerbate the effects on mental health.[47]

VII. Conclusion

In 2021, 4.3 billion individuals had a social media account, accounting for over half the world population and over 90 percent of internet users (We Are Social 2021). The repercussions of the rise of social media are thus likely to be far-reaching. In this paper, we leveraged the staggered introduction of Facebook across US colleges to estimate the impact of social media on mental health and found that the introduction of Facebook at a college had a negative effect on student mental health. Our evidence points to unfavorable social comparisons as the leading mechanism.

Overall, our results are consistent with the hypothesis that social media might be partly responsible for the recent deterioration in mental health among teenagers and young adults. It is up to social media platforms, regulators, and future research to determine whether and how these effects can be alleviated.

[46] The effects might be mitigated if, over time, users learn how to interpret the content they are exposed to on Facebook. The effects could be exacerbated if, over time, users become dependent on and potentially even addicted to Facebook (Allcott, Gentzkow, and Song 2021). A change in the social norms around the content that people post on social media might also affect the relationship between Facebook use and mental health.

[47] Of course, some of the changes underwent by social media platforms might push in the opposite direction. For instance, the increased popularity of Facebook might dilute the effects of social comparisons by changing the reference group from one's peers to a broader and more diverse set of individuals.

REFERENCES

Acquisti, Alessandro, and Ralph Gross. 2006. "Imagined Communities: Awareness, Information Sharing, and Privacy on the Facebook." In *Privacy Enhancing Technologies: 6th International Workshop, PET 2006, Cambridge, UK, June 2006, Revised Selected Papers*, edited by George Danezis and Philippe Golle, 36–58. New York: Springer.

►**Albert, Paul R.** 2015. "Why Is Depression More Prevalent in Women?" *Journal of Psychiatry & Neuroscience* 40 (4): 219–21.

►**Allcott, Hunt, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow.** 2020. "The Welfare Effects of Social Media." *American Economic Review* 110 (3): 629–76.

►**Allcott, Hunt, Matthew Gentzkow, and Lena Song.** 2022. "Digital Addiction." *American Economic Review* 112 (7): 2424–63.

American College Health Association (ACHA). 2000–2019. "American College Health Association-National College Health Assessment (ACHA-NCHA) Spring 2000–2019 Reference Group Data Report (Abridged)." *Journal of American College Health* 54 (4): 201–12.

American Psychological Association (APA). 2013. "Diagnostic and Statistical Manual of Mental Disorders: DSM-5."

►**Anderson, Michael L.** 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association* 103 (484): 1481–95.

►**Appel, Helmut, Alexander L. Gerlach, and Jan Crusius.** 2016. "The Interplay between Facebook Use, Social Comparison, Envy, and Depression." *Current Opinion in Psychology* 9: 44–49.

Aral, Sinan. 2021. *The Hype Machine: How Social Media Disrupts Our Elections, Our Economy, and Our Health–and How We Must Adapt.* New York: Currency.

Armona, Luis. 2019. "Online Social Network Effects in Labor Markets: Evidence from Facebook's Entry into College Campuses." SSRN 3381938.

Arrington, Michael. 2005. "85% of college students use Facebook." *TechCrunch*, February 4. https://techcrunch.com/2005/09/07/85-of-college-students-use-facebook.

►**Baranov, Victoria, Sonia Bhalotra, Pietro Biroli, and Joanna Maselko.** 2020. "Maternal Depression, Women's Empowerment, and Parental Investment: Evidence from a Randomized Controlled Trial." *American Economic Review* 110 (3): 824–59.

►**Bekalu, Mesfin A., Rachel F. McCloud, and K. Viswanath.** 2019. "Association of Social Media Use with Social Well-Being, Positive Mental Health, and Self-Rated Health: Disentangling Routine Use from Emotional Connection to Use." *Health Education and Behavior* 46 (S2): 69S–80S.

►**Benjamini, Yoav, Abba M. Krieger, and Daniel Yekutieli.** 2006. "Adaptive Linear Step-up Procedures That Control the False Discovery Rate." *Biometrika* 93 (3): 491–507.

►**Berryman, Chloe, Christopher J. Ferguson, and Charles Negy.** 2018. "Social Media Use and Mental Health among Young Adults." *Psychiatric Quarterly* 89 (2): 307–14.

Biasi, Barbara, Michael S. Dahl, and Petra Moser. 2019. "Career Effects of Mental Health." SSRN 2544251.

►**Blattman, Christopher, Julian C. Jamison, and Margaret Sheridan.** 2017. "Reducing Crime and Violence: Experimental Evidence from Cognitive Behavioral Therapy in Liberia." *American Economic Review* 107 (4): 1165–1206.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess. 2021. "Revisiting Event Study Designs: Robust and Efficient Estimation." Unpublished.

Braghieri, Luca, Ro'ee Levy, and Alexey Makarin. 2022. "Replication Data for: Social Media and Mental Health." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E175582V1.

Brügger, Niels. 2015. "A Brief History of Facebook as a Media Text: The Development of an Empty Structure." *First Monday* 20 (5).

Bursztyn, Leonardo, Georgy Egorov, Ruben Enikolopov, and Maria Petrova. 2019. "Social Media and Xenophobia: Evidence from Russia." Unpublished.

►**Callaway, Brantly, and Pedro H. C. Sant'Anna.** 2021. "Difference-in-differences with Multiple Time Periods." *Journal of Econometrics* 225 (2): 200–230.

Chan, David. 2010. "So Why Ask Me? Are Self-Report Data Really That Bad?" In *Statistical and Methodological Myths and Urban Legends*, edited by Charles E. Lance and Robert J. Vandenberg, 309–36. New York: Routledge.

►**De Chaisemartin, Clément, and Xavier d'Haultfoeuille.** 2020. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects." *American Economic Review* 110 (9): 2964–96.

►**Dienlin, Tobias, Philipp K. Masur, and Sabine Trepte.** 2017. "Reinforcement or displacement? The Reciprocity of FtF, IM, and SNS Communication and Their Effects on Loneliness and Life Satisfaction." *Journal of Computer-Mediated Communication* 22 (2): 71–87.

Donati, Dante, Ruben Durante, Francesco Sobbrio, and Dijana Zejcirovic. 2022. "Lost in the Net? Broadband Internet and Youth Mental Health." Unpublished.

►**Enikolopov, Ruben, Alexey Makarin, and Maria Petrova.** 2020. "Social Media and Protest Participation: Evidence from Russia." *Econometrica* 88 (4): 1479–514.

►**Enikolopov, Ruben, Maria Petrova, and Konstantin Sonin.** 2018. "Social Media and Corruption." *American Economic Journal: Applied Economics* 10 (1): 150–74.

Facebook. 2021. "Facebook Reports First Quarter 2021 Results." Facebook. https://bit.ly/3BBVfE3.

Fergusson, Leopoldo, and Carlos Molina. 2020. "Facebook Causes Protests." Unpublished.

Fujiwara, Thomas, Karsten Müller, and Carlo Schwarz. 2021. "The Effect of Social Media on Elections: Evidence from the United States." NBER Working Paper 28849.

►**Golberstein, Ezra, Gilbert Gonzales, and Ellen Meara.** 2019. "How Do Economic Downturns Affect the Mental Health of Children? Evidence from the National Health Interview Survey." *Health Economics* 28 (8): 955–70.

►**Goodman-Bacon, Andrew.** 2021. "Difference-in-differences with Variation in Treatment Timing." *Journal of Econometrics* 225 (2): 254–77.

Griffiths, Mark D., Daria J. Kuss, and Zsolt Demetrovics. 2014. "Social Networking Addiction: An Overview of Preliminary Findings." In *Behavioral Addictions*, edited by Kenneth Paul Rosenberg and Laura Curtiss Feder, 119–141. San Diego: Academic Press.

GWI. 2021. *Social GWI's Flagship Report on the Latest Trends in Social Media.* London: GWI.

Hass, Nancy. 2006. "In Your Facebook.com." *New York Times*. January 8. https://www.nytimes.com/2006/01/08/education/edlife/in-your-facebookcom.html.

Haushofer, Johannes, and Jeremy Shapiro. 2016. "The Short-Term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya." *Quarterly Journal of Economics* 132 (4): 2057–60.

Jacobs, Abigail Z., Samuel F. Way, Johan Ugander, and Aaron Clauset. 2015. "Assembling Thefacebook: Using Heterogeneity to Understand Online Social Network Assembly." In *Proceedings of the 7th Annual ACM Web Science Conference*, Article 18. New York: Association for Computing Machinery.

►**James, Spencer L., Degu Abate, Kalkidan Hassen Abate, Solomon M. Abay, Cristiana Abbafati, Nooshin Abbasi, Hedayat Abbastabar, Foad Abd-Allah, Jemal Abdela, Ahmed Abdelalim, et al.** 2018. "Global, Regional, and National Incidence, Prevalence, and Years Lived with Disability for 354 Diseases and Injuries for 195 Countries and Territories, 1990–2017: A Systematic Analysis for the Global Burden of Disease Study 2017." *Lancet* 392 (10159): 1789–858.

►**Kays, Kristina, Kathleen Gathercoal, and William Buhrow.** 2012. "Does Survey Format Influence Self-Disclosure on Sensitive Question Items?" *Computers in Human Behavior* 28 (1): 251–56.

►**Kelly, Yvonne, Afshin Zilanawala, Cara Booker, and Amanda Sacker.** 2019. "Social Media Use and Adolescent Mental Health: Findings from the UK Millennium Cohort Study." *EClinicalMedicine* 6: 59–68.

►**Kessler, Ronald C., G. Paul Amminger, Sergio Aguilar-Gaxiola, Jordi Alonso, Sing Lee, and T. Bedirhan Ustun.** 2007. "Age of Onset of Mental Disorders: A Review of Recent Literature." *Current Opinion in Psychiatry* 20 (4): 359–64.

Kirkpatrick, David. 2011. *The Facebook Effect: The Inside Story of the Company That Is Connecting the World.* New York: Simon and Schuster.

Kolek, Ethan A., and Daniel Saunders. 2008. "Online Disclosure: An Empirical Examination of Undergraduate Facebook Profiles." *Journal of Student Affairs Research and Practice* 45 (1): 1–25.

►**Kroenke, Kurt, and Robert L. Spitzer.** 2002. "The PHQ-9: A New Depression Diagnostic and Severity Measure." *Psychiatric Annals* 32 (9): 509–15.

►**Kroenke, Kurt, Robert L. Spitzer, and Janet B. W. Williams.** 2001. "The PHQ-9: Validity of a Brief Depression Severity Measure." *Journal of General Internal Medicine* 16 (9): 606–13.

Lampe, Cliff, Nicole B. Ellison, and Charles Steinfield. 2008. "Changes in Use and Perception of Facebook." In *CSCW '08: Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work*, edited by Bo Begole and David W. McDonald, 721–30. New York: Association for Computing Machinery.

Layard, Richard. 2017. "The Economics of Mental Health." *IZA World of Labor*: Article 321. https://doi.org/10.15185/izawol.321.

►**Levy, Ro'ee.** 2021. "Social Media, News Consumption, and Polarization: Evidence from a Field Experiment." *American Economic Review* 111 (3): 831–70.

►**Liu, Liu Yi, Jaime E. Sidani, Ariel Shensa, Ana Radovic, Elizabeth Miller, Jason B. Colditz, Beth L. Hoffman, Leila M. Giles, and Brian A. Primack.** 2016. "Association between Social Media Use and Depression among US Young Adults." *Depression and Anxiety* 33 (4): 323–31.

Meier, Adrian, Leonard Reinecke, and Christine E. Meltzer. 2016. "'Facebocrastination'? Predictors of using Facebook for Procrastination and Its Effects on Students' Well-Being." *Computers in Human Behavior* 64: 65–76.

►**Morgan, Elizabeth M., Chareen Snelson, and Patt Elison-Bowers.** 2010. "Image and Video Disclosure of Substance Use on Social Media Websites." *Computers in Human Behavior* 26 (6): 1405–11.

►**Mosquera, Roberto, Mofioluwasademi Odunowo, Trent McNamara, Xiongfei Guo, and Ragan Petrie.** 2020. "The Economic Effects of Facebook." *Experimental Economics* 23 (2): 575–602.

Müller, Karsten, and Carlo Schwarz. 2020. "From Hashtag to Hate Crime: Twitter and Anti-Minority Sentiment." SSRN 3149103.

National Alliance on Mental Illness (NAMI). 2020. *Mental Health by the Numbers.* Pittsburgh, PA: NAMI.

National Center for Health Statistics (NCHS). 2021. "Health, United States, 2019: Table 007." National Center for Health Statistics. https://www.cdc.gov/nchs/hus/contents2019.htm.

National Survey on Drug Use and Health (NSDUH). 2019. "2002–2019 National Survey on Drug Use and Health Final Analytic File (Codebook)." National Survey on Drug Use and Health. https://bit.ly/3e1d8Sg.

Nolen-Hoeksema, Susan, and Lori M. Hilt. 2008. "Gender Differences in Depression." In *Handbook of Depression*, edited by Ian H. Gotlib and Constance L. Hammen, 386–404. New York: Guilford Press.

►**Patel, Vikram, Alan J. Flisher, Sarah Hetrick, and Patrick McGorry.** 2007. "Mental Health of Young People: A Global Public-Health Challenge." *Lancet* 369 (9569): 1302–13.

►**Paul, Jomon Aliyas, Hope M. Baker, and Justin Daniel Cochran.** 2012. "Effect of Online Social Networking on Student Academic Performance." *Computers in Human Behavior* 28 (6): 2117–27.

►**Paul, Karsten I., and Klaus Moser.** 2009. "Unemployment Impairs Mental Health: Meta-Analyses." *Journal of Vocational Behavior* 74 (3): 264–82.

►**Persson, Petra, and Maya Rossin-Slater.** 2018. "Family Ruptures, Stress, and the Mental Health of the Next Generation." *American Economic Review* 108 (4-5): 1214–52.

►**Plomin, Robert, Claire M. A. Haworth, and Oliver S. P. Davis.** 2009. "Common Disorders are Quantitative Traits." *Nature Review Genetics* 10 (12): 872–78.

►**Puhl, Rebecca, and Young Suh.** 2015. "Stigma and Eating and Weight Disorders." *Current Psychiatry Reports* 17 (3): 1–10.

Ridley, Matthew, Gautam Rao, Frank Schilbach, and Vikram Patel. 2020. "Poverty, Depression, and Anxiety: Causal Evidence and Mechanisms." *Science* 370 (6522): 1289–1289. https://doi.org/10.1126/science.aay0214.

Roth, Jonathan, Pedro H. C. Sant'Anna, Alyssa Bilinski, and John Poe. 2022. "What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature." Unpublished.

►**Salk, Rachel H., Janet S. Hyde, and Lyn Y. Abramson.** 2017. "Gender Differences in Depression in Representative National Samples: Meta-Analyses of Diagnoses and Symptoms." *Psychological Bulletin* 143 (8): 783–822.

Shapiro, Bradley. 2021. "Promoting Wellness or Waste? Evidence from Antidepressant Advertising." *American Economic Journal: Microeconomics* 14 (2): 439–77.

►**Spitzer, Robert L., Kurt Kroenke, Janet B. W. Williams, and Bernd Löwe.** 2006. "A Brief Measure for Assessing Generalized Anxiety Disorder: The GAD-7." *Archives of Internal Medicine* 166 (10): 1092–97.

Stutzman, Frederic. 2006. "Student Life on the Facebook." Unpublished.

►**Sun, Liyang, and Sarah Abraham.** 2021. "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects." *Journal of Econometrics* 225 (2): 175–99.

Thompson, Sharon H., and Eric Lougheed. 2012. "Frazzled by Facebook? An Exploratory Study of Gender Differences in Social Network Communication among Undergraduate Men and Women." *College Student Journal* 46 (1): 88–98.

►**Traud, Amanda L., Peter J. Mucha, and Mason A. Porter.** 2012. "Social Structure of Facebook Networks." *Physica A: Statistical Mechanics and its Applications* 391 (16): 4165–80.

Twenge, Jean M. 2017. *iGen: Why Today's Super-Connected Kids Are Growing Up Less Rebellious, More Tolerant, Less Happy–and Completely Unprepared for Adulthood–and What That Means for the Rest of Us.* New York, NY: Simon and Schuster.

►**Twenge, Jean M., A. Bell Cooper, Thomas E. Joiner, Mary E. Duffy, and Sarah G. Binau.** 2019. "Age, Period, and Cohort Trends in Mood Disorder Indicators and Suicide-Related Outcomes in a Nationally Representative Dataset, 2005–2017." *Journal of Abnormal Psychology* 128 (3): 185–99.

►**Twenge, Jean M., and W. Keith Campbell.** 2019. "Media Use Is Linked to Lower Psychological Well-Being: Evidence from Three Datasets." *Psychiatric Quarterly* 90 (2): 311–31.

US Department of Education, National Center for Education Statistics. 2005. "Integrated Postsecondary Education Data System (IPEDS)." National Center for Education Statistics.

►**Verduyn, Philippe, Nino Gugushvili, Karlijn Massar, Karin Täht, and Ethan Kross.** 2020. "Social Comparisons on Social Networking Sites." *Current Opinion in Psychology* 36: 32–37.

►**Vogel, Erin A., Jason P. Rose, Lindsay R. Roberts, and Katheryn Eckles.** 2014. "Social Comparison, Social Media, and Self-Esteem." *Psychology of Popular Media Culture* 3 (4): 206–22.

Watson, Sandy White, Zachary Smith, and Jennifer Driver. 2006. "Alcohol, Sex and Illegal Activities: An Analysis of Selected Facebook Central Photos in Fifty States." Unpublished.

We Are Social. 2021. *Digital 2021 April Global Statshot Report*. New York: We Are Social.

World Health Organization. 2008. *The Global Burden of Disease: 2004 Update*. Geneva: World Health Organization.

►**Wilson, Robert E., Samuel D. Gosling, and Lindsay T. Graham.** 2012. "A Review of Facebook Research in the Social Sciences." *Perspectives on Psychological Science* 7 (3): 203–20.

►**Yang, Longfei, Yinghao Zhao, Yicun Wang, Lei Liu, Xiuyi Zhang, Bingjin Li, and Ranji Cui.** 2015. "The Effects of Psychological Stress on Depression." *Current Neuropharmacology* 13 (4): 494–504.

►**Zhuravskaya, Ekaterina, Maria Petrova, and Ruben Enikolopov.** 2020. "Political Effects of the Internet and Social Media." *Annual Review of Economics* 12: 415–38.

This article has been cited by:

1. Rui Guan, Hua Lei, Yan Zhi. 2025. Fertility and female mental health. *Journal of Asian Economics* **97**, 101862. [Crossref]
2. Liping Liao, Shuning Kong, Minzhe Du. 2025. The effect of clean heating policy on individual health: Evidence from China. *China Economic Review* **89**, 102309. [Crossref]
3. Yi Si, Minfeng Yu, Lei Zhang, Qing (Clara) Zhou. 2025. Board reforms and firm employment: Worldwide evidence. *Journal of Banking & Finance* **171**, 107379. [Crossref]
4. Mario Fernandes, Martin G. Becker, Kerstin Pull, Andreas Walter. 2025. When a B becomes an A: causal evidence on the effects of a journal ranking update on academics' publication behavior. *Studies in Higher Education* 1-22. [Crossref]
5. Peng Hou. 2025. Influence mechanism of Internet use on the physical and mental health of the Chinese elderly—Based on Chinese General Social Survey. *PLOS ONE* **20**:1, e0312664. [Crossref]
6. Jesus Barrena-Martinez, Maria-Jose Foncubierta-Rodriguez, José Luis Gilabert Villard. Emerging Psychosocial Risks in the Digital Age 151-182. [Crossref]
7. Jean M. Twenge, Dale Peeples, Reem M. A. Shafi, Elroy Boers, Lauren Hale, Paul Weigle. Social Media and Depressive Symptoms 137-140. [Crossref]
8. Ronen Gradwohl, Yuval Heller, Arye Hillman. 2025. How social media can undermine democracy. *European Journal of Political Economy* **86**, 102634. [Crossref]
9. Janet Currie, N. Meltem Daysal, Mette Gørtz, Jonas Lau-Jensen Hirani. 2025. Child Disability and Effects on Sibling Mental Health. *SSRN Electronic Journal* **59**. . [Crossref]
10. Begoña Cabeza Martinez, Beatrice d'Hombres, Matija Kovacic. Social Media Use, Loneliness and Emotional Distress Among Young People in Europe **24**, . [Crossref]
11. Emmanuel Bruce, Zhao Shurong, John Amoah, Sulemana Bankuoru Egala, Francis Kofi Sobre Frimpong. 2024. Reassessing the impact of social media on healthcare delivery: insights from a less digitalized economy. *Cogent Public Health* **11**:1. . [Crossref]
12. Raffaele Ciriello, Uri Gal, Oliver Hannon, Jason Thatcher. 2024. Responsible social media use: how user characteristics shape the actualisation of ambiguous affordances. *European Journal of Information Systems* **104**, 1-23. [Crossref]
13. Qinghu Jin, Sirui Wu. 2024. The impact of the classified voting system on corporate investment and equity value. *Contemporary Accounting Research* **84**. . [Crossref]
14. Balázs Lengyel, Gergő Tóth, Nicholas A. Christakis, Anikó Bíró. 2024. Antidepressant use in spatial social networks. *Science Advances* **10**:49. . [Crossref]
15. Guy Aridor, Rafael Jiménez-Durán, Ro'ee Levy, Lena Song. 2024. The Economics of Social Media. *Journal of Economic Literature* **62**:4, 1422-1474. [Abstract] [View PDF article] [PDF with links]
16. Fiona Scott Morton. 2024. Can Massive Technological Progress Hurt Workers? A Review of Power and Progress by Daron Acemoglu and Simon Johnson. *Journal of Economic Literature* **62**:4, 1671-1681. [Abstract] [View PDF article] [PDF with links]
17. Ramin Mojtabai. 2024. Problematic social media use and psychological symptoms in adolescents. *Social Psychiatry and Psychiatric Epidemiology* **59**:12, 2271-2278. [Crossref]
18. Peter Königs. 2024. In Defense of 'Surveillance Capitalism'. *Philosophy & Technology* **37**:4. . [Crossref]
19. Chun Xia, Jia Xu, Jiaxing Cheng, Yongmei Hou. 2024. The impact of information overload on the information avoidance of medical staff: The moderating and mediating roles of job performance and time pressure. *Computers in Human Behavior Reports* **16**, 100474. [Crossref]

20. Laura Marciano, Jeffrey Lin, Taisuke Sato, Sundas Saboor, Kasisomayajula Viswanath. 2024. Does social media use make us happy? A meta-analysis on social media and positive well-being outcomes. *SSM - Mental Health* **6**, 100331. [Crossref]

21. Sumer S. Vaid, Lara Kroencke, Mahnaz Roshanaei, Sanaz Talaifar, Jeffrey T. Hancock, Mitja D. Back, Samuel D. Gosling, Nilam Ram, Gabriella M. Harari. 2024. Variation in social media sensitivity across people and contexts. *Scientific Reports* **14**:1. . [Crossref]

22. Victoria Oldemburgo de Mello, Felix Cheung, Michael Inzlicht. 2024. Twitter (X) use predicts substantial changes in well-being, polarization, sense of belonging, and outrage. *Communications Psychology* **2**:1. . [Crossref]

23. Nick Obradovich, Sahib S. Khalsa, Waqas U. Khan, Jina Suh, Roy H. Perlis, Olusola Ajilore, Martin P. Paulus. 2024. Opportunities and risks of large language models in psychiatry. *NPP—Digital Psychiatry and Neuroscience* **2**:1. . [Crossref]

24. Renhui Fu, Fang Gao, Yi Zhao. 2024. The capital market consequences of stock market liberalisation: Evidence from Mainland-Hong Kong Stock Connect Programs in China. *Accounting & Finance* **64**:4, 3275-3299. [Crossref]

25. Samantha Marsh, Joanna Ting Wai Chu, Amanda Jane Calder. 2024. 'I tried to take my phone off my daughter, and i got hit in the face': a qualitative study of parents' challenges with adolescents' screen use and a toolbox of their tips. *BMC Public Health* **24**:1. . [Crossref]

26. Hege Ragnhildslakken, Tore Bonsaksen, Eivind Aakhus, Isaac Kabelenga, Gary Lamph, Daicia Price, Amy Østertun Geirdal. 2024. Social Media Use and Associations with Psychological Distress Among Older Adults During the COVID-19 Pandemic. *Social Sciences* **13**:12, 634. [Crossref]

27. Abhinita Daiya, Archit Garg. 2024. Thumbs up, ticket down: Decoding the disconnect between social media buzz and box office numbers. *International Journal of System Assurance Engineering and Management* **7**. . [Crossref]

28. Jianzeng An, Chun Xia, Jia Xu, Weiwei Li, Jianwen Ding. 2024. Dancing with care: promoting social integration and participation in community activities for older adults. *Frontiers in Public Health* **12**. . [Crossref]

29. Gilbert Gonzales, Tara McKay. 2024. Negative LGBTQ+ News Media Means Bad News for LGBTQ + Young Adult Mental Health. *JAMA Pediatrics* **178**:11, 1101. [Crossref]

30. Renhui Fu, Fang Gao, Yi Zhao. 2024. The impact of stock market liberalization on management forecast precision–Evidence from Mainland-Hong Kong Stock Connect Programs in China. *Journal of Accounting and Public Policy* **48**, 107243. [Crossref]

31. Yang Yu, Yituan Liu. 2024. Environmental regulations and firms' trade-offs between innovations: Empirical evidence from the quasi-experiment in China. *Journal of Environmental Management* **370**, 122819. [Crossref]

32. Dmitry Arkhangelsky, Guido Imbens. 2024. Causal models for longitudinal and panel data: a survey. *The Econometrics Journal* **27**:3, C1-C61. [Crossref]

33. Warren Cheng. 2024. Could Social Media Be a Means to Combat Mental Health Stigma?. *International Journal of Information Systems and Social Change* **15**:1, 1-15. [Crossref]

34. Clemente Pignatti, Zachary Parolin. 2024. The effects of an unconditional cash transfer on parents' mental health in the United States. *Health Economics* **33**:10, 2253-2287. [Crossref]

35. Jane Cooley Fruehwirth, Alex Xingbang Weng, Krista M. Perreira. 2024. The effect of social media use on mental health of college students during the pandemic. *Health Economics* **33**:10, 2229-2252. [Crossref]

36. N. Liu, S. Wu, J. Guo. 2024. Spillover health losses of COVID-19 in China: evidence from non-COVID-19-related mortality. *Public Health* **235**, 8-14. [Crossref]

37. Kevin Arceneaux, Martial Foucault, Kalli Giannelos, Jonathan Ladd, Can Zengin. 2024. Facebook increases political knowledge, reduces well-being and informational treatments do little to help. *Royal Society Open Science* **11**:10. . [Crossref]

38. Patric Plesa. 2024. Neonihilism: Meaninglessness and irony in neoliberal capitalism. *Theory & Psychology* **34**:5, 611-629. [Crossref]

39. Brittany Hannouch, Tema Milstein. 2024. Activating Ecocentrism: How Young Women Environmental Activists Produce Identity on Instagram. *Environmental Communication* **6**, 1-20. [Crossref]

40. Rui Zhang, Dawei Feng, Ruifeng Teng. 2024. Network infrastructure and individual mental health: evidence from China. *Applied Economics* 1-15. [Crossref]

41. Jiaping Zhang, Xiaomei Gong. 2024. How mobile social networks affect rural household income: evidence from China. *Kybernetes* **130**. . [Crossref]

42. Tanusree Chakraborty, V. S. Sharada, Daisy Gohain. Harmony Within 95-130. [Crossref]

43. Simona-Vasilica Oprea, Adela Bâra. 2024. Assessing the Dual Impact of the Social Media Platforms on Psychological Well-being: A Multiple-Option Descriptive-Predictive Framework. *Computational Economics* . [Crossref]

44. Tae-Hyun Kim. 2024. The Impact of a Coastal Water Quality Policy in South Korea: Evidence from the Total Pollution Load Management Program. *Water Economics and Policy* **10**:03. . [Crossref]

45. David Lazer, Briony Swire-Thompson, Christo Wilson. 2024. A Normative Framework for Assessing the Information Curation Algorithms of the Internet. *Perspectives on Psychological Science* **19**:5, 749-757. [Crossref]

46. Jennifer K Wagner, Megan Doerr, Cason D Schmit. 2024. AI Governance: A Challenge for Public Health. *JMIR Public Health and Surveillance* **10**, e58358-e58358. [Crossref]

47. Ngawang Dendup, Dil Bahadur Rahut, Yayue Xiao. 2024. Short-run impact of electricity on social capital: evidence from a rural electricity program. *Empirical Economics* **114**. . [Crossref]

48. Zhilong Zhao, Lin Zhu, Jing Liao, Jiaxin Xia, Xueya Pu. 2024. Parental Mediation, Digital Media Usage, and Health Literacy: An Exploration Among Chinese Elementary School Students. *Health Communication* 1-13. [Crossref]

49. Xiyu Yuan. Social Media Sentiment Analysis and Mental Health Prediction using Deep Belief Network (DBN) 1-5. [Crossref]

50. José Ramón Enríquez, Horacio Larreguy, John Marshall, Alberto Simpser. 2024. Mass Political Information on Social Media: Facebook Ads, Electorate Saturation, and Electoral Accountability in Mexico. *Journal of the European Economic Association* **22**:4, 1678-1722. [Crossref]

51. Patrick Fagan. 2024. Clicks and tricks: The dark art of online persuasion. *Current Opinion in Psychology* **58**, 101844. [Crossref]

52. Shilan Feng, Bingrui Gao, Ya Tan, Kaiyuan Xian, Yingjia Zhai. 2024. Resource-based transformation and urban resilience promotion: Evidence from firms' carbon emissions reductions in China. *Journal of Cleaner Production* **468**, 143118. [Crossref]

53. Benjamin Cowan, Todd R. Jones, Jeffrey Swigert. 2024. Parental and Student Time Use Around the Academic Year. *Journal of Economic Behavior & Organization* **224**, 66-110. [Crossref]

54. Sunday Azagba, Todd Ebling, Alperen Korkmaz. 2024. Social media use and mental health indicators among US adolescents: A population-based study. *Journal of Psychiatric Research* **176**, 354-359. [Crossref]

55. Ailyn D. Diaz, Dale Aaron Peeples, Paul Weigle. 2024. Depression and Social Media Use in Children and Adolescents. *Pediatric Clinics of North America* **24**. . [Crossref]

56. Dandan Xu, Shiqi Feng, Zijin Huang. 2024. Positive and negative factors of the influence of new media and the digital environment on ideal and belief education among university students in China. *The Journal of General Psychology* **14**, 1-24. [Crossref]
57. Maurizio Pugno. 2024. Social media effects on well-being: The hypothesis of addiction of a new variety. *Kyklos* **77**:3, 690-704. [Crossref]
58. Hang Zheng, Jian Han, Yiqun Liu. 2024. Authorized economic operator certification and export stability: Evidence from China's firms. *Review of International Economics* **32**:3, 1174-1203. [Crossref]
59. Monika Haga, Vilde Malmin, Ove Østerlie. 2024. Secondary school students' perception of health and teaching of health dimensions in physical education. *Physical Education and Sport Pedagogy* 1-13. [Crossref]
60. Chhabi Lal Ranabhat, Jason W Marion. Use of social media increases the risk of anxiety depression globally: results from 113 countries **30**, . [Crossref]
61. Yiwei Liu, Qiuyue Guo. 2024. Impact of Adult Children's Socioeconomic Status on the Health of Older Adults in China. *Journal of Family and Economic Issues* **132**. . [Crossref]
62. Akshay Ajit, Crispin Mathew, David Johns Denny, Devanath H, Dhanya P M. Mental State Report From Social Media Activity 1-5. [Crossref]
63. Ishfaq Hussain Bhat, Shilpi Gupta, Ghulam Mohammad Bhat. 2024. Effect of social media usage on major depressive disorder among generation Z: a study in Indian context. *Information Discovery and Delivery* **52**:3, 261-272. [Crossref]
64. Oupin Tang, Xiaomeng Shi, Lili Jiu. 2024. Value creation or political trick? An event study on anti-ESG regulations. *Finance Research Letters* **65**, 105530. [Crossref]
65. Rui Ma, Fei Guo, Dongdong Li. 2024. Can public data availability affect stock price crash risk? Evidence from China. *International Review of Financial Analysis* **94**, 103270. [Crossref]
66. Jintong Tang, Wenping Ye, Mingzhi Hu, Stephen X. Zhang, Shaji A. Khan. 2024. The gendered effect of populism on innovation. *Journal of Business Venturing* **39**:4, 106393. [Crossref]
67. Zhenggang Wang, Zenan Wu, Ye Yuan. 2024. We've got you covered! The effect of public health insurance on rural entrepreneurship in China. *Journal of Public Economics* **235**, 105150. [Crossref]
68. Rakesh Allu, Maya Ganesh, Sarang Deo, Sripad K. Devalkar. 2024. Technology-Enabled Agent Choice and Uptake of Social Assistance Programs: Evidence from India's Food Security Program. *Manufacturing & Service Operations Management* **26**:4, 1472-1488. [Crossref]
69. Nearchos Paspallis, Panayiotis Panayiotou. An Assessment of ML-based Sentiment Analysis for Intelligent Web Filtering 80-87. [Crossref]
70. Charlene Marie Kalenkoski, Sabrina Wulff Pabilonia. 2024. Teen social interactions and well-being during the COVID-19 pandemic. *Review of Economics of the Household* **37**. . [Crossref]
71. Phoebe Chin, Faye Gorman, Fraser Beck, Bruce R. Russell, Klaas E. Stephan, Olivia K. Harrison. 2024. A systematic review of brief respiratory, embodiment, cognitive, and mindfulness interventions to reduce state anxiety. *Frontiers in Psychology* **15**. . [Crossref]
72. Rhea Milson, Daniel J. Madigan. 2024. Picture Perfect. *Journal of Media Psychology* **17**. . [Crossref]
73. Ceren Budak, Brendan Nyhan, David M. Rothschild, Emily Thorson, Duncan J. Watts. 2024. Misunderstanding the harms of online misinformation. *Nature* **630**:8015, 45-53. [Crossref]
74. Keqi Huang, Qiren Liu. 2024. Artificial intelligence and outward foreign direct investment: Evidence from China. *Journal of Asian Economics* **92**, 101745. [Crossref]
75. G.S. Prakasha, Sangeetha Rangasamy, Vaddi Jahnavitha, Sarah Dlima. 2024. Students' usage of Over-the-top (OTT) streaming platforms affecting their academic and socio-demographic profile. *Innoeduca. International Journal of Technology and Educational Innovation* **10**:1, 124-143. [Crossref]

76. Fabio Sabatini. 2024. The behavioral, economic, and political impact of the Internet and social media: Empirical challenges and approaches. *Journal of Economic Surveys* . [Crossref]

77. Angelina Voggenreiter, Sophie Brandt, Fabian Putterer, Andreas Frings, Juergen Pfeffer. The Role of Likes: How Online Feedback Impacts Users' Mental Health 302-310. [Crossref]

78. George Beknazar-Yuzbashev, Rafael Jiménez-Durán, Mateusz Stalinski. 2024. A Model of Harmful Yet Engaging Content on Social Media. *AEA Papers and Proceedings* **114**, 678-683. [Abstract] [View PDF article] [PDF with links]

79. Qiuyu Mao. 2024. Digitalization in prefabricated constructions and reducing carbon emission: empirical evidence from China. *Environmental Science and Pollution Research* **31**:22, 32914-32934. [Crossref]

80. Ryo Takahashi, Kengo Igei, Yusuke Tsugawa, Makiko Nakamuro. 2024. The effect of silent eating during school lunchtime on COVID-19 outbreaks. *Social Science & Medicine* **348**, 116852. [Crossref]

81. Michael L. Platt, Peter Sterling. 2024. Declining human fertility and the epidemic of despair. *Nature Mental Health* **2**:5, 463-465. [Crossref]

82. Preben Bertelsen, Simon Ozer. 2024. A good enough grip on life in youth. *Nordic Psychology* **5**, 1-25. [Crossref]

83. Andrew C Pickett, Danny Valdez, Kelsey L Sinclair, Wesley J Kochell, Boone Fowler, Nicole E Werner. Social Media Discourse Related to Caregiving for Older Adults Living With Alzheimer Disease and Related Dementias: Computational and Qualitative Study (Preprint) . [Crossref]

84. Binhai Wei, Minghong Luo, Chunkai Zhao. 2024. Bridging the gap: the effect of rural e-commerce development on internal income inequality in China. *Applied Economics* 1-20. [Crossref]

85. Xin Su, Shengwen Wang. 2024. The impact and management of internet-based public opinion dissemination during emergencies: A case study of Baidu News during the first wave of coronavirus disease 2019 (COVID-19). *PLOS ONE* **19**:4, e0299374. [Crossref]

86. Martin Hilbert, Arti Thakur, Pablo M. Flores, Xiaoya Zhang, Jee Young Bhan, Patrick Bernhard, Feng Ji. 2024. 8–10% of algorithmic recommendations are 'bad', but... an exploratory risk-utility meta-analysis and its regulatory implications. *International Journal of Information Management* **75**, 102743. [Crossref]

87. Elena Fumagalli, L. J. Shrum, Tina M. Lowrey. 2024. The Effects of Social Media Consumption on Adolescent Psychological Well-Being. *Journal of the Association for Consumer Research* **9**:2, 119-130. [Crossref]

88. Andranik Tumasjan. 2024. The many faces of social media in business and economics research: Taking stock of the literature and looking into the future. *Journal of Economic Surveys* **38**:2, 389–426. [Crossref]

89. Marula Bizjak Ferjan, Jernej Buzeti, Janez Stare, Špela Mar. Zdravje in dobro počutje študentov 91-102. [Crossref]

90. Shengnan Ma. 2024. Compulsory education and gender gaps in intrahousehold empowerment. *Review of Economics of the Household* **100**. . [Crossref]

91. Sparshi Srivastava, Jamie T. Mullins. 2024. Temperature, Mental Health, and Individual Crises: Evidence from Crisis Text Line. *American Journal of Health Economics* **103**. . [Crossref]

92. Dave E. Marcotte, Benjamin Hansen. 2024. The re-emerging suicide crisis in the U.S.: Patterns, causes and solutions. *Journal of Policy Analysis and Management* **43**:2, 582-612. [Crossref]

93. Yu Wu, Eva Kemps, Ivanka Prichard. 2024. Digging into digital buffets: A systematic review of eating-related social media content and its relationship with body image and eating behaviours. *Body Image* **48**, 101650. [Crossref]

94. Cameron J. Bunker, Julia M. Balcerowska, Lena-Marie Precht, Jürgen Margraf, Julia Brailovskaia. 2024. Perceiving the self as authentic on social media precedes fewer mental health symptoms: A longitudinal approach. *Computers in Human Behavior* **152**, 108056. [Crossref]
95. Weimin He, Bin Wang. 2024. Environmental jurisdiction and energy efficiency: Evidence from China's establishment of environmental courts. *Energy Economics* **131**, 107358. [Crossref]
96. Jianping Liu, Zixue Tai, Fengbin Hu. 2024. Prevalence and coping of depression and anxiety among college students during COVID-19 lockdowns in China. *Journal of Affective Disorders* **348**, 305-313. [Crossref]
97. Thomas Potrebny, Sondre Aasen Nilsen, Anders Bakken, Tilmann von Soest, Kirsti Kvaløy, Oddrun Samdal, Børge Sivertsen, Heidi Aase, Lasse Bang. 2024. Secular trends in mental health problems among young people in Norway: a review and meta-analysis. *European Child & Adolescent Psychiatry* **48**. . [Crossref]
98. Kate Lister, Elena Riva, Alison Hartley, Philippa Waterhouse, Naomi Moller, Leigh Downes, Tim Coughlan, Agnes Kukulska-Hulme, Elaine McPherson, Ian Macdonald, Sophie Jones-Tinsley, Cath Brown, Ruth Tudor. 2024. Positive Digital Practices: Supporting Positive Learner Identities and Student Mental Wellbeing in Technology-Enhanced Higher Education. *Journal of Interactive Media in Education* **2024**:1. . [Crossref]
99. Jie Chen, Tapas Mishra, Wei Song, Qingjing Zhang, Zhuang Zhang. 2024. The impact of bank mergers on corporate tax aggressiveness. *Journal of Corporate Finance* **84**, 102540. [Crossref]
100. Lijuan Lv, Yan Chen. 2024. The Collision of digital and green: Digital transformation and green economic efficiency. *Journal of Environmental Management* **351**, 119906. [Crossref]
101. Sara Ayllón, Pablo Brugarolas, Samuel Lado. ICT Use and Children's Self-reported Life Satisfaction 239-279. [Crossref]
102. Béatrice d'Hombres, Chiara Gentile. Social Media Use and Loneliness 93-115. [Crossref]
103. Abdullah Al Mamun, Shahrur Anonna Akhy, Sumaya Mustofa, Md. Badol Mia, Partha Dip Sarkar, Narayan Ranjan Chakraborty, Md. Abbas Ali Khan. Impact of the Use of Social Media Among University Students Using Machine Learning 219-231. [Crossref]
104. Maurizio Pugno. 2024. Creativity, well-being, and economic development: An evolutionary approach. *Journal of Evolutionary Economics* **34**:1, 205-225. [Crossref]
105. Sybren Slimmen, Olaf Timmermans, Lilian Lechner, Anke Oenema. 2024. A socio-ecological approach of evidence on associations between social environmental factors and mental health outcomes of young adults: A systematic review. *Social Sciences & Humanities Open* **10**, 101068. [Crossref]
106. Yunita Furi Aristyasari, Abdul Mustaqim, Ichsan, Chusnul Azhar. 2024. Implementation of Psychotherapeutic in Islamic Education. *SHS Web of Conferences* **204**, 05002. [Crossref]
107. Alireza Pourkeyvan, Ramin Safa, Ali Sorourkhah. 2024. Harnessing the Power of Hugging Face Transformers for Predicting Mental Health Disorders in Social Networks. *IEEE Access* **12**, 28025-28035. [Crossref]
108. Brandon T. McDaniel, Jenny Radesky, Jessica Pater, Adam M. Galovan, Annaliese Harrison, Victor Cornet, Lauren Reining, Alexandria Schaller, Michelle Drouin. 2024. Heavy Users, Mobile Gamers, and Social Networkers: Patterns of Objective Smartphone Use in Parents of Infants and Associations With Parent Depression, Sleep, Parenting, and Problematic Phone Use. *Human Behavior and Emerging Technologies* **2024**:1. . [Crossref]
109. Qiuyu Mao. Digitalization in Prefabricated Constructions and Reducing Carbon Emission: Empirical Evidence from China **47**, . [Crossref]
110. Fabio Sabatini. 2024. The Behavioral, Economic, and Political Impact of the Internet and Social Media: Empirical Challenges and Approaches. *SSRN Electronic Journal* **15**. . [Crossref]

111. Guy Aridor, Rafael Jiménez Durán, Ro'ee Levy, Lena Song. 2024. The Economics of Social Media. *SSRN Electronic Journal* **14**. . [Crossref]

112. Guy Aridor, Rafael Jiménez-Durán, Ro'ee Levy, Lena Song. 2024. The Economics of Social Media. *SSRN Electronic Journal* **14**. . [Crossref]

113. Vanisha Sharma. 2024. Social (Media) Learning: Experimental Evidence from Indian Farmers. *SSRN Electronic Journal* **2**. . [Crossref]

114. Tse-Chun Lin, Yican Liu, Fangzhou Lu. 2024. Happy App Happy Tip: The Return Predictability from Amusement Apps Downloads Around the World. *SSRN Electronic Journal* **122**. . [Crossref]

115. Lei Zhang, Yi Si, Minfeng Yu, Qing (Clara) Zhou. Board Reforms and Firm Employment: Worldwide Evidence **33**, . [Crossref]

116. Namgyoon Oh. 2024. Nutrition to Nurturance: The Impact of Children's WIC Eligibility Loss on Parental Well-being. *SSRN Electronic Journal* **97**. . [Crossref]

117. Sumit Agarwal, Mohit Desai, Pulak Ghosh, Nishant Vats. 2024. Bridging the Information Gap: Sowing the Seeds of Productivity with High-Speed 4G Internet. *SSRN Electronic Journal* **24**. . [Crossref]

118. J. Anthony Cookson, William Mullins, Marina Niessner. 2024. Social Media and Finance. *SSRN Electronic Journal* **118**. . [Crossref]

119. Yinan Liu, Xianhua Zai. Do Safe Assets Affect the Financial Decisions of Individuals Nearing Retirement? Evidence from the Partnership for Long-Term Care Program **86**, . [Crossref]

120. Osea Giuntella, Silvia Saccardo, Sally Sadoff. 2024. Sleep: Educational Impact and Habit Formation. *SSRN Electronic Journal* **6**. . [Crossref]

121. Felix Jan Nitsch, Klaus Wertenbroch, Hilke Plassmann. Why People Post **39**, . [Crossref]

122. Qianyun Gao, Yu Gao, Juanjuan Meng, Shizhe Yu. Algorithmic Personalization and Digital Addiction: A Field Experiment on Douyin (TikTok) **101**, . [Crossref]

123. Daron Acemoglu, Daniel Huttenlocher, Asuman E. Ozdaglar, James Siderius. 2024. Online Business Models, Digital Ads, and User Welfare. *SSRN Electronic Journal* **14**. . [Crossref]

124. Chenchen Li, Frank Zhang, Kailiang Zhang. For-benefit or For-profit? The Dark Side of Stakeholderism Legislation **27**, . [Crossref]

125. Andrew Leigh, Stephen Robson. The Rise of Social Media and the Fall in Mental Wellbeing Among Young Australians **112**, . [Crossref]

126. Wei Jia, Lei Liu, Gang Peng. 2024. The Impact of Social Media on Users' Self-Efficacy and Loneliness: An Analysis of the Mediating Mechanism of Social Support. *Psychology Research and Behavior Management* **Volume 17**, 593-612. [Crossref]

127. Liping Fu, Caiping Liu, Yongqing Dong, Xiaodong Ma, Quanling Cai, Dongli Li, Kaisheng Di. 2024. Mediating Effects of Information Access on Internet Use and Multidimensional Health Among Middle-Aged and Older Adults: Nationwide Cross-Sectional Study. *Journal of Medical Internet Research* **26**, e49688. [Crossref]

128. Andrew C Pickett, Danny Valdez, Kelsey L Sinclair, Wesley J Kochell, Boone Fowler, Nicole E Werner. 2024. Social Media Discourse Related to Caregiving for Older Adults Living With Alzheimer Disease and Related Dementias: Computational and Qualitative Study. *JMIR Aging* **7**, e59294. [Crossref]

129. Zhiping Jia. 2024. Quantitative Impact Analysis of Social Network Usage Patterns on College Students' Mental Health and Preventive Countermeasures. *Applied Mathematics and Nonlinear Sciences* **9**:1. . [Crossref]

130. Ruguo Song, Wenjing Li, Tingting Li, Xiaoling Liu. 2024. Enhancing collaborative learning in psychoeducation through social network analysis. *Applied Mathematics and Nonlinear Sciences* 9:1. . [Crossref]

131. Yangyang Cheng, Zhenhuan Xu. 2024. Fiscal policy promotes corporate green credit: Experience from the construction of energy conservation and emission reduction demonstration cities in China. *Green Finance* 6:1, 1-23. [Crossref]

132. Tajim Md. Niamat Ullah Akhund, Deep Ajabani, Zaffar Ahmed Shaikh, Ali Elrashidi, Waleed A. Nureldeen, Muhammad Ishaq Bhatti, M Mesbahuddin Sarker. 2024. A comprehensive exploration of human communal media interaction and its evolving impact on psychological health across demographics and time. *PeerJ Computer Science* **10**, e2398. [Crossref]

133. Yao Wang, Chunchao Wang, Rui Zhang. 2023. Parental non-cognitive abilities and child mental health. *Journal of Asian Economics* **89**, 101667. [Crossref]

134. Xiangyuan Ding, Luoqi Yuan, Yi Zhou. 2023. Internet access and older adults' health: Evidence from China. *China Economic Review* **82**, 102047. [Crossref]

135. Donghui Xie, Caiquan Bai, Yuwei Zhang. 2023. Relation-based governance, financial crisis shock, and economic growth in China. *Economic Modelling* **129**, 106565. [Crossref]

136. Yangyang Cheng, Zhenhuan Xu. 2023. Fiscal centralization and urban industrial pollution emissions reduction: Evidence from the vertical reform of environmental administrations in China. *Journal of Environmental Management* **347**, 119212. [Crossref]

137. Rimsha Mahrukh, Sadia Shakil, Aamir Saeed Malik. 2023. Sentiments analysis of fMRI using automatically generated stimuli labels under naturalistic paradigm. *Scientific Reports* **13**:1. . [Crossref]

138. Akihiro Nishi, Christopher A. German, Sage K. Iwamoto, Nicholas A. Christakis. 2023. Status invisibility alleviates the economic gradient in happiness in social network experiments. *Nature Mental Health* **1**:12, 990-1000. [Crossref]

139. Luke Balcombe. 2023. AI Chatbots in Digital Mental Health. *Informatics* **10**:4, 82. [Crossref]

140. Cristopher Martin Olivares Guido, Claudia Amelia Diaz Olavarrieta. 2023. GROOMING Y SEXTING: LO QUE NECESITAS SABER. *Boletín Sapiens* :7, 1-7. [Crossref]

141. Jovana Dudukovic, Souliyany Gnotsyoudom, Oleksandra Trysna, Yazan Mahmoud Qandeel, Andrea Tick. The Psychological Impact of TikTok on Mental Health 000213-000218. [Crossref]

142. Luis H. Favela, Mary Jean Amon. The ethics of human digital twins: Counterfeit people, personhood, and the right to privacy 1-6. [Crossref]

143. Mario Fernandes, Simon Hilber, Jan-Egbert Sturm, Andreas Walter. 2023. Closing the gender gap in academia? Evidence from an affirmative action program. *Research Policy* **52**:9, 104865. [Crossref]

144. Dong Zhou, Yi Xu, Qixuan He. 2023. New Media Use and Mental Health of Married Women: Mediating Effects of Marital Quality. *Healthcare* **11**:21, 2909. [Crossref]

145. Jing Zhao, Xufeng Zhu. 2023. Spreading expertise: think tanks as digital advocators in the social media era. *Policy and Society* **42**:3, 359-377. [Crossref]

146. Prachi H. Bhuptani, Gabriela López, Roselyn Peterson, Lindsay M. Orchowski. 2023. Online Social Reactions to Disclosure of Sexual Victimization via #MeToo and Symptoms of Post-Traumatic Stress Disorder. *Journal of Interpersonal Violence* **38**:19-20, 10900-10919. [Crossref]

147. Sarah K. Competiello, George Y. Bizer, D. Catherine Walker. 2023. The Power of Social Media: Stigmatizing Content Affects Perceptions of Mental Health Care. *Social Media + Society* **9**:4. . [Crossref]

148. Ayala Arad, Ohad Barlizaly, Maayan Perchick. 2023. Facebook, social comparison and happiness: Evidence from a quasi-natural experiment. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace* **17**:4. . [Crossref]

149. Alexandru Savu. 2023. Temperature highs, climate change salience, and Eco-anxiety: early evidence from the 2022 United Kingdom heatwave. *Applied Economics Letters* 1-8. [Crossref]

150. Tanjim Mahmud, Michal Ptaszynski, Juuso Eronen, Fumito Masui. 2023. Cyberbullying detection for low-resource languages and dialects: Review of the state of the art. *Information Processing & Management* **60**:3, 103454. [Crossref]

151. Lauren Eales, Andrea Wiglesworth, Bonnie Klimes-Dougan, Kathryn R. Cullen. 2023. Editorial: Benign Versus Problematic Media Use Profiles in Youth: Correlates, Concerns, and Future Directions. *Journal of the American Academy of Child & Adolescent Psychiatry* **62**:9, 970-972. [Crossref]

152. Tim Donkers, Jürgen Ziegler. 2023. De-sounding echo chambers: Simulation-based analysis of polarization dynamics in social networks. *Online Social Networks and Media* **37-38**, 100275. [Crossref]

153. Alireza Pourkeyvan, Ramin Safa, Ali Sorourkhah. Harnessing the Power of Hugging Face Transformers for Predicting Mental Health Disorders in Social Networks **370**, . [Crossref]

154. Tingting Xie, Ye Yuan, Hui Zhang. 2023. Information, awareness, and mental health: Evidence from air pollution disclosure in China. *Journal of Environmental Economics and Management* **120**, 102827. [Crossref]

155. Martin Salzmann-Erikson, Henrik Eriksson. 2023. A Mapping Review of Netnography in Nursing. *Qualitative Health Research* **33**:8-9, 701-714. [Crossref]

156. Keqi Huang, Qiren Liu, Chengjian Tang. 2023. Which firms benefit from robot adoption? Evidence from China. *Journal of Asian Economics* **86**, 101612. [Crossref]

157. Sergio Currarini, Francesco Feri, Bjoern Hartig, Miguel A. Meléndez-Jiménez. 2023. Transmission and use of information in network games. *European Economic Review* **155**, 104443. [Crossref]

158. Tobias Schultheiss, Curdin Pfister, Ann-Sophie Gnehm, Uschi Backes-Gellner. 2023. Education expansion and high-skill job opportunities for workers: Does a rising tide lift all boats?. *Labour Economics* **82**, 102354. [Crossref]

159. Henry Onyeaka, Joseph Firth, Kobi Vanessa Ajayi, Chioma Muoghalu, Karima Holmes, Stanley Nkemjika, Funso Adeolu, Gibson Anugwom, Precious Obehi Eseaton, Nelson C. Onyeaka, Jeff Huffman, John Torous. 2023. Association between social media use and health promotion among individuals with depression and anxiety: Insights from the 2017–2020 Health Information National Trends Survey. *Journal of Mood and Anxiety Disorders* **1**, 100006. [Crossref]

160. Joelle Gross, Samuel Baltz, Charles Stewart. 2023. What Happens When the President Calls You an "Enemy of the People?" Election Officials and Public Sentiment. *Election Law Journal: Rules, Politics, and Policy* **22**:2, 185-211. [Crossref]

161. Claes Ek, Margaret Samahita. 2023. Too much commitment? An online experiment with tempting YouTube content. *Journal of Economic Behavior & Organization* **208**, 21-38. [Crossref]

162. Kelton Minor, Kristoffer Lind Glavind, Aaron J. Schwartz, Christopher M. Danforth, Sune Lehmann, Andreas Bjerre-Nielsen. 2023. Nature Exposure is Associated With Reduced Smartphone Use. *Environment and Behavior* **55**:3, 103-139. [Crossref]

163. Climent Quintana-Domeque, Jingya Zeng, Xiaohui Zhang. 2023. Internet and mental health during the COVID-19 pandemic: evidence from the UK. *Oxford Open Economics* **2**. . [Crossref]

164. Tisha Sadariya, Shanti Verma. Early Prediction and Detection of Anxiety Level Using Support Vector Machine 279-291. [Crossref]

165. Avani K. Patel, Simone A. Bernstein, Jessica A. Gold. A Nuanced Approach to Adolescent Mental Health in the Era of Social Media . [Crossref]

166. Ning Liu, Hong Qian, Ben Zhong Zhang, Jing Guo. 2023. Spillover effects of violent attacks and COVID-19 exposure on mental health of health professionals: A two-phase quasi-natural experiments study in Northwest China. *Cambridge Prisms: Global Mental Health* **10**. . [Crossref]

167. Suppawong Tuarob, Krittin Chatrinan, Thanapon Noraset, Tanisa Tawichsri, Tipajin Thaipisutikul. 2023. Forecasting National-Level Self-Harm Trends With Social Networks. *IEEE Access* **11**, 64796-64814. [Crossref]

168. Martin Hilbert, Arti Thakur, Feng Ji, Pablo M. Flores, Xiaoya Zhang, Jee Young Bhan, Patrick Bernhard. 2023. Recommender Algorithms Do No Harm -90% But... An Exploratory Risk-Utility Meta-Analysis of Algorithmic Audits. *SSRN Electronic Journal* **1**. . [Crossref]

169. Garrett Johnson, Tesary Lin, James C. Cooper, Liang Zhong. 2023. COPPAcalypse? The Youtube Settlement's Impact on Kids Content. *SSRN Electronic Journal* **54**. . [Crossref]

170. Clemente Pignatti, Zachary Parolin. 2023. The Effects of an Unconditional Cash Transfer on Mental Health in the United States. *SSRN Electronic Journal* **38**. . [Crossref]

171. Vikas Gawai. 2023. Broadband Technology, Aging, and Mental Health. *SSRN Electronic Journal* **112**. . [Crossref]

172. Rui Ma, Fei Guo, Dongdong Li. Can Public Data Availability Affect Stock Price Crash Risk? Evidence from China **100**, . [Crossref]

173. Guy Axel Ruiz Morales. 2023. Mental Health and Mobile Internet Access. Evidence from Mexico. *SSRN Electronic Journal* **9**. . [Crossref]

174. Jin Liu, Xingchen (Cedric) Xu, Yongjun Li, Yong Tan. 2023. 'Generate' the Future of Work through AI: Empirical Evidence from Online Labor Markets. *SSRN Electronic Journal* **40**. . [Crossref]

175. Leonardo Bursztyn, Benjamin Handel, Rafael Jiménez-Durán, Christopher Roth. 2023. When Product Markets Become Collective Traps: The Case of Social Media. *SSRN Electronic Journal* **87**. . [Crossref]

176. You Hong. 2023. Does Destination Cities' Digitalization Block Vulnerable Groups from Moving to Opportunity: Evidence from China. *SSRN Electronic Journal* **113**. . [Crossref]

177. Sergei Guriev, Emeric Henry, Théo Marquis, Ekaterina Zhuravskaya. 2023. Curtailing False News, Amplifying Truth. *SSRN Electronic Journal* **70**. . [Crossref]

178. Yoshito Funashima. 2023. Media-Created Economic Uncertainty. *SSRN Electronic Journal* **57**. . [Crossref]

179. Leonardo Bursztyn, Benjamin Handel, Rafael Jiménez-Durán, Christopher Roth. 2023. When Product Markets Become Collective Traps: The Case of Social Media. *SSRN Electronic Journal* **87**. . [Crossref]

180. Hong You. Does Destination Cities' Digitalization Block Vulnerable Groups from Moving to Opportunity: Evidence from China **138**, . [Crossref]

181. Mukesh Eswaran. 2023. Suffering, Identity, and the Economics of Religion: Implications for Subjective Wellbeing and the Good Life. *SSRN Electronic Journal* **105**. . [Crossref]

182. Samina Ashraf, Idnan Adrees, RuKhsana Bashir. 2022. Perceptions of Hearing-impaired Students on the Role of Social Media for their Inclusion into Mainstream Society. *Journal of Business and Social Review in Emerging Economies* **8**:2. . [Crossref]

183. Dante Donati, Ruben Durante, Francesco Sobbrio, Dijana Zejcirovic. 2022. Lost in the Net? Broadband Internet and Youth Mental Health. *SSRN Electronic Journal* **6**. . [Crossref]

184. Rafael Jiménez Durán. 2022. The Economics of Content Moderation: Theory and Experimental Evidence from Hate Speech on Twitter. *SSRN Electronic Journal* **110**. . [Crossref]

185. Dante Donati, Ruben Durante, Francesco Sobbrio, Dijana Zejcirovic. 2022. Lost in the Net? Broadband Internet and Youth Mental Health. *SSRN Electronic Journal* **6**. . [Crossref]

186. Berkeren Buyukeren, Alexey Makarin, Heyu Xiong. 2022. The Impact of Dating Apps on Young Adults: Evidence From Tinder. *SSRN Electronic Journal* **61**. . [Crossref]

187. Rakesh Allu, Maya Ganesh, Sarang Deo, Sripad K. Devalkar. 2022. Agent-Choice in Last-Mile Delivery of Food Security Programs: Impact, Usage and Implications. *SSRN Electronic Journal* **9**. . [Crossref]

188. Raphaela Andres, Olga Slivko. 2021. Combating Online Hate Speech: The Impact of Legislation on Twitter. *SSRN Electronic Journal* **110**. . [Crossref]

189. Julia Cage, Nicolas Herve, Beatrice Mazoyer. 2020. Social Media and Newsroom Production Decisions. *SSRN Electronic Journal* **10**. . [Crossref]

10

Psychiatry Research 328 (2023) 115496

Contents lists available at ScienceDirect

Psychiatry Research

journal homepage: www.elsevier.com/locate/psychres

Youth screening depression: Validation of the Patient Health Questionnaire-9 (PHQ-9) in a representative sample of adolescents

Eduardo Fonseca-Pedrero[a,*], Adriana Díez-Gómez[a], Alicia Pérez-Albéniz[a], Susana Al-Halabi[b], Beatriz Lucas-Molina[c], Martin Debbané[d,e]

[a] Universidad de La Rioja, Spain
[b] Universidad de Oviedo, Spain
[c] Universidad de Valencia, Spain
[d] Developmental Clinical Psychology Research Unit, Faculty of Psychology and Educational Sciences, University of Geneva, Switzerland
[e] Research Department of Clinical, Educational and Health Psychology, University College London, United Kingdom

ARTICLE INFO

Keywords:
Depression
Adolescents
Psychometric
Validation
PHQ-9
Mental health

ABSTRACT

Background: Depression symptoms and mood disorders constitute one of the major public health challenges among youths. Thus, early prevention and intervention for depression should be a priority. The main goal of the present study was to validate the Patient Health Questionnaire-9 (PHQ-9) scores in a school-based sample of non-clinical adolescents.

Method: Stratified random sampling was conducted. Participants were 2235 students (M = 14.49, SD =1.76, range= 12–18 years), 52.9 % were female, from 34 secondary schools in Spain. Several previously validated self-reported questionnaires of mental health and psychopathology were administrated.

Results: The unidimensional factorial model of the PHQ-9 items showed adequate goodness of fit indices. Strong measurement invariance across gender was found. Omega for the PHQ-9 total score was 0.87. The PHQ-9 total score was positively associated with anxiety symptoms and emotional and behavioral problems, and negatively associated with prosocial behavior and quality of life.

Conclusions: The PHQ-9 is a brief, easy, and reliable tool for assessing self-reported depressive symptoms in both clinical and school settings. PHQ-9 may be used as a screening tool for universal early detection and monitorization of depression symptoms during adolescence.

1. Introduction

Emotional problems (e.g., depression, anxiety) are among the leading causes of associated disability and global burden of disease in young people. The *Global Burden of Diseases, Injuries, and Risk Factors Study* GBD (2019) (GBD 2019 Mental Disorders Collaborators, 2022) revealed that mental disorders remained among the top ten causes of burden of disease worldwide, with no evidence of an overall reduction in burden since 1990. The World Health Organization (WHO, 2022) informs that one billion people worldwide have a diagnosis of a mental disorder (more than one in eight adults and adolescents). Depression in young people is a rising concern, so it is our duty as a society to promote, protect and care for the mental health of the entire population, in particular of one of the most vulnerable groups, children and adolescents. Thus, missed opportunities for depression identification and treatment can be quite costly from both personal and public health perspectives (Davis et al., 2022).

Depression, as a continuum phenotype, encompasses a range of mood-related concepts and a spectrum of difficulties, placing at its most extreme end clinical disorders that could be expressed as mood syndromes; while at the other end, depression can refer to a mood state in the context of normative affective experience (Hankin, 2015; Thapar et al., 2022). Across this quantitative variation, intermediate (subclinical or subthreshold depression) expressions can also be identified. Subthreshold depression is associated with low mood and additional depressive symptoms such as loss of interest and enjoyment, but without reaching the diagnostic threshold (Thapar et al., 2022). Previous works have found that youth depression is continuously, not categorically, distributed. This viewpoint of the dimensional alternative to traditional Nosologies (e.g., Diagnostic and Statistical Manual of Mental Disorders)

* Corresponding author at: University of La Rioja, C/ Luis de Ulloa, 2, Edificio VIVES; C.P: 26004, Logroño, La Rioja, Spain.
E-mail address: eduardo.fonseca@unirioja.es (E. Fonseca-Pedrero).

https://doi.org/10.1016/j.psychres.2023.115496
Received 17 July 2023; Received in revised form 14 September 2023; Accepted 15 September 2023
Available online 16 September 2023

are related with psychopathology models as, the Hierarchical Taxonomy of Psychopathology (HiTOP) (Kotov et al., 2021). Conceptualizing depression (internalizing spectra) as a graded dimension has implications for theory and practice, as well as for methods and measurement (Ruiz et al., 2023; Hankin et al., 2005).

The prevalence put forward by developmental epidemiological research for major depressive disorder and dysthymia is situated around 8 % (95 % CI: 0.02–0.13) and 4 % (95 % CI: 0.01–0.07) among adolescents, respectively. In addition, the global point prevalence rate of elevated self-reported depressive symptoms from 2001 to 2020 was 34 % (95 % CI: 0.30–.38) (Shorey et al., 2022). That is, 34 % of adolescents worldwide, aged 10–19 years, are at risk of developing clinical depression, which exceeds the reported estimates of individuals aged 18 to 25 years. The age of onset, severity, persistence, and comorbidity are factors to consider in the study of depression symptoms during this developmental stage. Previous studies have found that the average age of development of any mental disorders is 14.5 years (Solmi et al., 2022). The proportion of individuals with onset of mood disorders before the age of 14, 18, 25 were, respectively, 2.5 %, 11.5 %, 34.5 %, and the peak age was 20.5 years (k = 79, median = 31, IQR = 21–46) (Solmi et al., 2022). Additionally, adverse outcomes associated with clinical and subclinical depression during adolescence include the onset of other mental health disorders (e.g., anxiety, substance abuse, and conduct disorders). Elevated depressive symptoms are associated with many outcomes as risk behaviours, health problems, and adverse psychosocial outcomes in interpersonal, social, educational, and occupational functioning as well as suicidal behaviours (Thapar et al., 2022). For instance, in the educational sphere, depression was associated with poorer school grades (Riglin et al., 2014) and school absenteeism (Finning et al., 2019).

Depression symptoms and disorders in adolescents are frequently misdiagnosed and undertreated. Routine screening, from a preventive approach, has the potential to improve the early and reliable identification of depressive symptoms. According to NICE Guidelines [NG134] for depression in children and young people, healthcare professionals in primary care, schools and other relevant community settings should be trained to detect symptoms of depression, and to assess children and young people who may be at risk of depression (NICE Guidance, 2019). The detection of these individuals with subclinical depression, whether in health, social or educational settings, requires the availability of adequate tools to make informed and data-driven decisions. The PHQ-9 (Kroenke et al., 2001; Spitzer et al., 1999) has become a standard measure of depression research and clinical practice. The PHQ-9 is a self-report developed to assess the severity of depression according to DSM criteria. Its psychometric properties have been adequately examined (El-Den et al., 2018; Kroenke, 2021). Previous studies showed strong evidence that the PHQ-9 can be used as a unidimensional measure of depressive symptoms (Bianchi et al., 2022). In primary care settings, the brief nature and ease of scoring of this instrument make it an excellent choice for providers and researchers looking to implement depression screening (Richardson et al., 2010). The evidence suggests that assessing the factors separately will not provide any useful information for most patients (Boothroyd et al., 2019). An adolescent version of the PHQ-9 was also designed (Johnson et al., 2002). In addition, the standard PHQ-9 has been validated in previous work with adolescents (Anum et al., 2019; Borghero et al., 2018; Burdzovic & Brunborg, 2017; Fatiregun and Kumapayi, 2014; Leung et al., 2020; Khow et al., 2016; Richardson et al., 2010; Sinclair-McBride et al., 2018). For instance, Burdzovic and Brunborg (2017), in a sample of Norwegian adolescents, found a single-factor structure for the PHQ-9.

Many instruments can be used to assess depression symptoms, but further adaptation of tests is needed to identify these experiences at an early age (Demaras et al., 2019). This developmental period involves physical, psychological, and social changes, which may increase an individual's sensitivity and reactivity to stress exposure (Bock et al., 2021). To date, although the standard PHQ-9 has been validated in previous studies, yet there is little information on the psychometric properties of the PHQ-9 scores in large and representative samples of the general population. For instance, few studies have tested the measurement invariance by gender or gathered new validity evidence using modern psychometric methods as item response theory (IRT) for this measurement tool.

In this context, the main goal of the present study was to validate the PHQ-9 scores in a school-based sample of adolescents. This study aimed to: a) analyze the prevalence of depressive symptoms; b) examine the internal structure of the PHQ-9 scores; c) test the measurement invariance of the PHQ-9 by gender; d) estimate the reliability of the PHQ-9 scores; and e) analyze the association between PHQ-9 scores and psychometric indicators of mental health and quality of life. In line with previous literature, it was hypothesized that the one-factor model of the PHQ-9 would have adequate goodness-of-fit indices. In addition, we further hypothesized that this hypothesized dimensional model would be equal across gender. We also expected that the reliability estimation of the PHQ-9 scores would be adequate. Finally, we expected that depressive symptoms would be related with emotional and behavioural difficulties.

2. Method

2.1. Participants

Stratified random sampling was conducted at the class level in the total student population of La Rioja (region in Northern Spain). The students belonged to different public and charter educational centres, compulsory secondary education and vocational training. Strata were formed depending on the public and charter nature of the educational institutions, and the educational level. A total of 34 schools and 98 classrooms participated in the study.

The initial sample consisted of 2640 students. Those participants that: a) showed a high score on the Oviedo Infrequency Response Scale (more than 2 points) (n = 175) and were over 18 years of age (n = 247) were removed. Thus, a total of 2235 students, 1045 men (46.8 %), 1183 (52.9 %) women, and 7 (0.3 %) non-binary identity participated in the study. The mean age was 14.49 years (SD=1.76), age range between 12 and 18 years. The age distribution was as follows: 12 years, n = 280; 13 years, n = 387; 14 years, n = 396; 15 years, n = 408; 16 years, n = 371; 17 years, n = 240; and 18 years, n = 153. The 90.8 % of the sample was identified as Spanish.

2.2. Instruments

Socio-demographics, mental health problems, and lifestyle. An ad hoc instrument was developed to assess age, school grade, sex, gender, and nationality. In addition, family history of mental disorder was assessed. Information about lifestyle was also collected: general health, hours of sleep, time to fall asleep, and frequency of free-time activities.

Patient Health Questionnaire-9 (PHQ-9) (Kroenke et al., 2001; Spitzer et al., 1999). The PHQ-9 is composed of nine questions designed to assess depressive symptomatology according to DSM criteria. The items are answered according to the frequency of symptoms (0 = *not at all*, 1= *some days*, 2 = *more than half of the days*, 3 = *almost every day*). A higher score is indicative of greater depressive symptomatology. The PHQ-9 has been validated into Spanish (González-Blanch et al., 2018).

Generalized Anxiety Disorder Assessment (GAD-7) (Spitzer et al., 2006). The GAD-7 is a seven-item instrument used to measure or assess the severity of generalized anxiety disorder. Each item asks the individual to rate the severity of his or her symptoms over the past two weeks. Response options include 0 = *not at all*, 1 = *several days*, 2 = *more than half the days*, 3 = *almost every day*. The GAD-7 has been validated into Spanish (Muñoz-Navarro et al., 2017). In this study, the reliability of the total score was adequate (McDonald's Omega = 0.90)

Strengths and Difficulties Questionnaire (SDQ) (Goodman, 1997). The

SDQ is a self-report questionnaire that is widely used for the assessment of different emotional and behavioural problems related to mental health in adolescents. The SDQ is made up of a total of 25 statements distributed across five subscales: Emotional symptoms, Conduct problems, Hyperactivity, Peer problems, and Prosocial behavior. The first four subscales yield a Total difficulties score. In this study we used a Likert-type response format with three options (0 = *not true*, 1 = *somewhat true*, 2 = *certainly true*). The validated Spanish version of the SDQ was used in the present study (Ortuño-Sierra et al., 2022). The SDQ total difficulties score showed adequate reliability in this sample (McDonald's Omega = 0.75).

Kidscreen-10 Index (Ravens-Sieberer et al., 2010). The Kidscreeen-10 Index is a measurement instrument developed and validated to assess health-related quality of life in children and adolescents aged 8 to 18 years. It presents a total of 10 questions in a Likert 5-choice response format, where a higher score is indicative of higher quality of life. The Kidscreen-10 has been validated in Spain (Aymerich et al., 2005). The Kidscreeen-10 scores showed good reliability in this sample (McDonald's Omega = 0.91).

The Oviedo Infrequency Scale-revised (INF-OV-R) (Fonseca-Pedrero et al., 2009). The INF-OV-R was administered to the participants to detect those who responded in a random, pseudorandom or dishonest manner. The INF-OV-R instrument is a self-report composed of 10 items in a dichotomic scale format (*Yes/No*). Students with more than two incorrect responses on the INF-OV-R scale were eliminated from the sample.

2.3. *Procedure*

The research was approved by the Ethical Committee of Clinical Research of La Rioja (CEImLAR, PI 552). The psychometric measures were administered collectively, through personal computers, in groups of 10 to 30 students, during school hours and in a classroom specially prepared for this purpose. Administration took place under the supervision of the researchers trained in a standard protocol. No incentive was provided for their participation. Participants' parents were asked to sign an informed consent form so that their children could participate in the study. Participants were informed of the confidentiality of their responses and of the voluntary nature of the study. This work is part of a broader project called PSICE (Evidence-based Psychology in Educational Contexts) (Fonseca-Pedrero et al., 2023b) (ClinicalTrials.gov. Ref. NCT05322642).

2.4. *Data analyses*

First, we calculated the prevalence and descriptive statistics of the PHQ-9 items. The PHQ-9 total score was divided into the following categories of increasing severity: 0–4 (minimal), 5–9 (mild), 10–14 (moderate), 15–19 (moderately severe), and 20–27 points (severe).

Second, in order to analyse the internal structure of the PHQ-9, several confirmatory factor analyses (CFA) were performed. Attending to previous studies, a one-dimensional model was examined. Diagonally Weighted Least Squares estimator was used. The following goodness-of-fit indices were used: Chi-square (χ^2), Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA) and 90 % Confidence interval, and Standardized Root Mean Square Residual (SRMR). Hu and Bentler (1999) suggested that RMSEA should be 0.06 or less for a good model fit and CFI and TLI should be 0.95 or more, though any value over 0.90 tends to be considered acceptable.

Third, in order to test measurement invariance across gender, successive multigroup CFAs were conducted. Basically, a hierarchical set of steps are followed when testing measurement invariance, typically starting with the determination of a well-fitting multigroup and baseline model and continuing with the establishment of successive equivalence constraints in the model parameters across groups. The baseline model is called the configural model, which is the first and least restrictive model specified and is important because it represents the baseline model against which all subsequent specified invariance models are compared. The configural model is established by specifying and testing the CFA model for each group separately. Once the theoretical model has been validated in both groups, configural invariance is then examined, requiring that the same pattern of fixed and freely estimated model parameters is equivalent across groups; however, no equality constraints are imposed on the model parameters between groups. Configural invariance is tested by assessing the model fit. When configural invariance is met (i.e., the model fits the data), it suggests that at least the general factor structure is similar, but not necessarily equivalent, across groups. The next step is to impose equality constraints on the factor loadings across the groups to test metric or weak invariance. If the model fit with the constrained parameters is significantly and practically worse than the baseline or configural model, then weak invariance is not supported. When metric invariance is met, it suggests that the same unit of measurement is being used for the item across the groups and that the participants interpret and respond to the measure in a similar manner (Horn and McArdle, 1992). The final step is to impose constraints on the item intercepts and factor loadings to test strong or scalar invariance across groups. The confirmation of the invariance of the intercepts permits comparison of the latent means in both groups. The analyzed models are nested in that the imposed constraints are progressively added. Due to the limitations of the $\Delta\ \chi 2$ regarding its sensitivity to sample size, Cheung and Rensvold (2002) proposed a more practical criterion, the Δ CFI, to determine if nested models are practically equivalent. In this study, when Δ CFI is greater than 0.01 between two nested models, the more constrained model is rejected since the additional constraints have produced practically worse fit. However, if the change in CFI is less than or equal to 0.01, it is considered that all specified equal constraints are tenable; therefore, we can continue with the next step in the analysis of measurement invariance.

Fourth, reliability estimation of the PHQ-9 scores were estimated using McDonald's Omega. In addition, from the IRT framework with the 2-PL Model, the test information function was computed. Classical test theory methods cannot give us direct guidance on the latent trait of a measure to accurately assess depressive experiences at various points along the continuum (Hambleton et al., 1991). IRT methods provide estimates of the position on the latent trait (i.e., the theta level) where the tool provides the most information. Test information function graphically depict the regions of the latent trait continuum most precisely assessed. Greater information reflects greater measurement accuracy, or reliability. Test information function are estimated on the same latent trait scale (standardized $M = 0$; $SD = 1$), yielding information that is comparable across tests (Olino et al., 2012).

Fifth, the associations between PHQ-9 scores and other mental health indicators were calculated. SPSS 22.0, FACTOR 10.5.01, and JASP were used for data analyses.

3. Results

3.1. *Descriptive statistics*

Prevalence and descriptive statistics for the PQH-9 items are shown in Table 1. The prevalence rates of depressive symptoms according to the recommended cut-off points were: 52.1 % (minimal), 20.2 % (mild), 15.2 % (moderate), 8.1 % (moderately severe), and 4.4 % (severe).

3.2. *Confirmatory factor analysis of the PHQ-9 items*

The standardized factor loadings for the total sample and by gender are shown in Table 2. Goodness-of-fit indices for the one-dimensional model were adequate (see Table 3).

Table 1
Prevalence (%) of response and descriptive statistics for the items of the Patient Health Questionnaire-9 (PHQ-9) in the whole sample.

Item	Not at all	Several days	More than half of the days	Almost every day	M	SD	Skewness	Kurtosis
1 Little interest or pleasure in doing things	37.3	44.4	10.5	7.8	0.89	0.88	0.90	0.21
2 Feeling down, depressed, or hopeless	36.2	42.9	11.9	9	0.94	0.91	0.83	−0.06
3 Trouble falling or staying asleep, or sleeping too much	53.1	25.4	9.4	12.1	0.80	1.03	1.05	−0.18
4 Feeling tired or having little energy	56.1	23.3	10.3	9.4	0.70	0.99	1.31	0.20
5 Poor appetite or overeating	27.7	44.7	14.4	13.2	1.13	0.97	0.62	−0.53
6 Feeling bad about yourself – or that you are a failure	56.5	26.4	8.5	8.5	0.69	0.95	1.26	0.51
7 Trouble concentrating on things	37.9	34.8	15.2	12	1.01	1.01	0.68	−0.64
8 Moving or speaking so slowly that other people could have noticed	63.8	22.4	9.3	4.6	0.55	0.84	1.47	1.26
9 Thoughts that you would be better off dead or of hurting yourself	82.5	11.2	3.5	2.7	0.26	0.65	2.75	7.26

Table 2
Standardized factor loadings of the Patient Health Questionnaire-9 (PHQ-9) for the total sample and by gender.

Items	Total sample	Male	Female
1	0.843	0.802	0.832
2	0.737	0.667	0.796
3	0.710	0.593	0.714
4	0.732	0.609	0.729
5	0.722	0.625	0.735
6	0.835	0.802	0.803
7	0.607	0.580	0.627
8	0.561	0.498	0.554
9	0.799	0.778	0.773

Note. All standardized factor loadings estimated were statistically significant ($p < .01$).

3.3. *Measurement invariance of the PHQ-9 scores across gender*

Since the one-factor model showed a good fit, the measurement invariance of the unidimensional model of the PHQ-9 was tested as a function of gender. Goodness-of-fit indices for males and females were adequate (see Table 3). The configural, metric invariance and scalar measurement invariance models showed an adequate fit to the data. The ΔCFI between the constrained and unconstrained models was under 0.01, thus measurement invariance across gender for this unidimensional model was supported.

3.4. *Reliability estimation of the PHQ-9 scores*

The internal consistency of the PHQ-9 total frequency score, estimated with McDonald's Omega, was 0.87 (95 % CI: 0.86–0.88). Item discrimination indices were higher than 0.30. According to the IRT framework, the test information function provides an optimal estimation at the medium-high latent trait (values between 0 and 2) (see Fig. 1). The tool reduces its accuracy around the lowest level of the latent trait.

3.5. *Evidence based on the relations of the PHQ-9 scores to other variables*

We also studied the correlation between the PHQ-9 total score and mental health psychometric indicators. As shown in Table 4, the PHQ-9 total score was positively and statistically significant correlated with anxiety symptoms, and emotional and behavioral problems, and negatively associated with prosocial behavior and quality of life.

4. Discussion

Depression symptoms and disorder is a common problem among adolescents (Shorey et al., 2022). Routine screening for depression is recommended, yet standardization in screening and management is lacking (Cucci et al., 2022). Improving early recognition and assessment procedures, using reliable tools, can help us to prevent, as well as to promote evidence-based psychological treatments for mild and

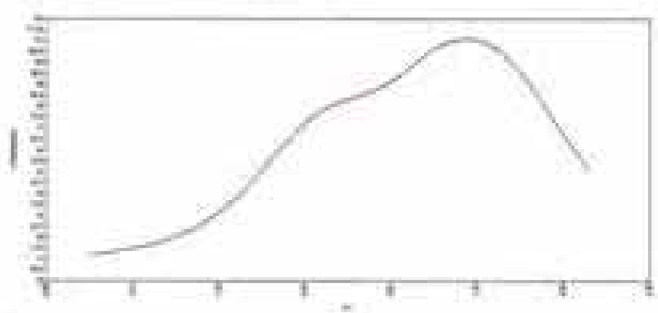


Fig. 1. Test information function of the Patient Health Questionnaire-9 (PHQ-9).
Note. F1=latent trait. Test information function graphically depict the regions of the latent trait continuum most precisely assessed. Greater information reflects greater measurement accuracy, or reliability.

Table 3
Goodness-of-fit indices for the hypothetical models tested and measurement invariance of the Patient Health Questionnaire-9 (PHQ-9) by gender.

Model	χ^2	*df*	CFI	TLI	RMSEA (90 % IC)	SRMR	ΔCFI[a]
One-factor	203.802	27	0.993	0.991	0.054 (0.054–0.061)	0.043	
Measurement invariance							
Male (n = 1045)	99.404	27	0.986	0.983	0.051 (0.040–0.062)	0.054	
Female (n = 1180)	131.934	27	0.993	0.991	0.057 (0.048–0.067)	0.046	
Configural invariance	230.832	54	0.992	0.989	0.054 (0.047–0.067)	0.050	
Metric invariance	275.388	62	0.990	0.988	0.056 (0.049–0.062)	0.054	<0.01
Scalar invariance	396.395	79	0.985	0.987	0.060 (0.054–0.066)	0.052	<0.01

Note.
χ^2 = Chi square; *df* = degrees of freedom; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = Root Mean Square Error of Approximation; IC = Interval Confidence; SRMR = Standardized Root Mean Square Residual; ΔCFI = Change in Comparative Fit Index.
Good model fit is indicated by a RMSEA ≤ 0.06, CFI and TLI ≥ 0.90 of 0.95, SRMR ≥ 0.08.
[a]CFI Δ < 0.01 indicates measurement invariance across gender.

Table 4
Pearson's correlations between the Patient Health Questionnaire-9 (PHQ-9) and the measures of mental health and quality of life.

	PHQ-9	GAD-7	SDQ PREM	SDQ PRCD	SDQ PRPP	SDQ HIP	SDQ PROS
GAD-7	0.766**						
SDQ PREM	0.709**	0.751**					
SDQ PRCD	0.397**	0.389**	0.391**				
SDQ PP	0.413**	0.383**	0.396**	0.257**			
SDQ HIP	0.474**	0.453**	0.362**	0.446**	0.184**		
SDQ PROS	−0.117**	−0.044*	0.004	−0.204**	−0.214**	−0.141**	
KS-10	−0.552**	−0.471**	−0.485**	−0.248**	−0.316**	−0.300**	0.177**

Note. **$p < .01$;
PHQ-9= Patient Health Questionnaire-9; GAD-7 = Generalized Anxiety Disorder Assessment-7; SDQ = Strengths and Difficulties Questionnaire; SDQ PREM = SDQ emotional problems; SDQ PRCD = SDQ conduct problems; SDQ PRPP = SDQ peer problems; SDQ HIP = SDQ hyperactivity; SDQ PROS = SDQ prosocial behavior; KS-10= KIDSCREEN-10 Index.

moderate to severe depression. Thus, the main objective of the present study was to validate the PHQ-9 scores in a community-based sample of adolescents. The prevention, using reliable screening methods, are fundamental in the management of this phenomena (Patra and Kumar, 2023), considered as one of the major barriers for the family, educational, health and societal systems.

During the last two weeks, the 12.5 % of the adolescent sample reported, symptoms of depression of moderately severe (8.1 %) and severe (4.4 %). The results found in the present study seem consistent with previous international reports examining self-reported depression symptoms in adolescent school samples using the PHQ-9 (Burdzovic and Brunborg, 2017; Tsai et al., 2014). For instance, 5.8 % prevalence of clinically-elevated symptoms among Norwegian adolescents or 5.1 %, among Chinese high-school students (Tsai et al., 2014). Nonetheless, since we used internationally defined cut-off scores but not tested in Spain, these results should be considered as preliminary. Furthermore, the PHQ-9 has a frequency-based item response system, which does not necessarily imply depressive severity.

The unidimensional factorial model of the PHQ-9 showed adequate goodness of fit indices. Similar results have been found in previous research (Burdzovic and Brunborg, 2017; Anum et al., 2019; Leung et al., 2020). For instance, Leung et al. (2020) found, using CFA, that the one-factor model with three pairs of item correlations fitted the PHQ-9 data well, and that measurement invariances by age and gender were supported. In another work, Burdzovic and Brunborg (2017) found a single-factor structure for the PHQ-9 and adequate reliability estimation for both genders. Also, these results are in line with those found in the adult population (e.g., Bianchi et al., 2022). At this regard, previous studies supported robust evidence for the unidimensional structure of PHQ-9 to assess depressive symptoms (Bianchi et al., 2022), and add useful information for most patients (Boothroyd et al., 2019). From a conceptual point of view and given the overlap with other emotional problems and disorders (e.g., anxiety), we should consider, during adolescence, the PHQ-9 total score as a proxy for the general dimension of emotional dysregulation or internalizing factor (e.g., Piqueras et al., 2021). This view is fully congruent with current transdiagnostic and psychopathology models (e.g., HiTOP) consider that psychological phenomena as representing unbroken spectra ranging from very low to very high levels (Kotov et al., 2023). From this broader framework, subjective distress should be largely determined by the presence of emotional -or internalizing symptoms- such as anxiety and depression (Piqueras et al., 2021).

Multi-group CFAs showed that the one-factor model of the PHQ-9 had strong measurement invariance across gender. Previous studies have found mixed results. For instance, Burdzovic & Brunborg (2017) found no evidence of metric or scalar equality across genders. Anum et al. (2019) conducted a multi-group CFA and supported a one-factor structure of the PHQ-9 that was invariant across gender. These results showed that all PHQ-9 items were equivalent across gender (none showed differential items functioning). It should be stressed that if measurement invariance does not hold, the validity of such scores should be questioned. Comparability between different groups only makes sense if it can be guaranteed that participants interpret and understand the latent construct in a similar manner (Horn and McArdle, 1992).

The reliability, estimated with McDonald's Omega, for the PHQ-9 total score was 0.87. This result is convergent with those found in previous studies conducted in adolescent populations (Burdzovic & Brunborg, 2017; Anum et al., 2019; Leung et al., 2020). In addition, we computed, from IRT, the test information function (TIF). The TIF provides an optimal estimation at the medium-high latent trait, that is, the PHQ-9 provided information at middle and higher levels of depression severity continuum. It is important in the context of ability or true score estimation because the TIF serves as an estimate of the latent trait accuracies of depression (i.e., the levels of the latent trait are measured with less standard error of measurement).

The PHQ-9 was positively associated with anxiety symptoms and emotional and behavioral problems, and negatively associated with prosocial behavior and quality of life. Similar results have found in prior research in both adolescent and adult samples. For example, Anum et al. (2019) found that the PHQ-9 correlated significantly with measures of anxiety, depression, mental wellbeing, and suicidal behavior. Furthermore, youth depression has been associated with a wide variety of risk and protective factors (e.g., Beck et al., 2021; Thapar et al., 2022). The risk factors identified include, among others, female sex, older age, poorer performance at school, lower interpersonal trust, social stress, atypicality, anxiety, feelings of incompetence, somatization, exposure to adverse events such as illness or death of a family member, physical or sexual abuse, bullying, poor academic achievement, poor sleep, more negative body image, more problematic use of social media or computer games, as well as poorer family functioning and inconsistent parental discipline (Beck et al., 2021).

One line of research in the field of depression is based on the idea of early detection, prevention and intervention in individuals who report subclinical depression with the aim of mitigating or reducing the impact that the disorder may have on the personal, family, academic, health and social spheres (Bernaras et al., 2019). Compared to other methods (e.g., clinical interviews), the use of these tools constitutes a rapid, efficient and non-invasive method of assessment. Moreover, the delay in diagnosis and treatment, the inadequate supply of mental health services, and the adverse consequences of depressive symptoms and disorders reinforce the importance of screening and treatment of this phenomena (as an internalizing disorder, depression is much less likely to be detected) at critical stages of human development such as adolescence. Due to the rising trend of depressive symptoms and the increasing number of at-risk teenagers, clinicians, researchers and practitioners should be more vigilant and proactive in their outreach activities to raise awareness and promote access and availability of services for this vulnerable group (Shorey et al., 2022). The United States Preventive Services Task Force recommends screening for depression in adolescents aged 12 to 18 years in the primary care setting (Grade B recommendation) (Forman-Hoffman et al., 2016).

Depression screening and prevention programs for young people need to go beyond the clinic walls (Guo and Zhu, 2021). Educational settings are at the forefront of mental health promotion and prevention during childhood and adolescence. Universal school-based screening would improve the detection of adolescent depression and reduce the limited capacity of the healthcare system to provide adequate mental health care to those who clearly need treatment. Schools are the "natural" place for actions to promote mental well-being and, specifically, for the prevention of mental health problems. Most adolescents spend long periods of time in classrooms, with schools being one of the main agents involved in socialization, as well as in formation and promotion of optimal development (Fonseca-Pedrero et al., 2023a). In this regard, the WHO Guidelines on School Health Services (WHO, 2021) highlight that schools are essential environments for the acquisition of socio-emotional skills (e.g., self-regulation and resilience). In the last decade, a range of psychological interventions to promote mental health and prevent mental health problems in schools have been tested with varying degrees of success (González-Roz et al., 2023).

Taken together, the results showed that the PHQ-9 is a brief, easy, and reliable tool for assessing self-reported depressive symptoms in school settings. This research provides further support for the validity of the PHQ-9 scores in a large and representative sample of non-clinical adolescents. In addition, the results indicated that the PHQ-9 could be a brief and useful tool to assess depression in general population samples. Future studies should continue to analyze the protective and risk factors for depression, add new psychometric procedures (e.g., network models) and digital based technologies (Elosua et al., 2023) to prevent mental health disorders in young people.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was funded by a national project awarded by the Ministry of Science and Innovation of the Government of Spain and the Agency and the European Regional Development Fund (Project "PID2021-127301OB-I00" funded by MCIN / AEI /10.13039/501100011033 FEDER, UE).

References

Anum, A., Adjorlolo, S., Kugbey, N., 2019. Depressive symptomatology in adolescents in Ghana: examination of psychometric properties of the Patient Health Questionnaire-9. J. Affect. Disord. 256, 213–218. https://doi.org/10.1016/j.jad.2019.06.007.
Aymerich, M., Berra, S., Guillamón, I., Herdman, M., Alonso, J., Ravens-Sieberer, U., Rajmil, L., 2005. Development of the Spanish version of the KIDSCREEN, a health-related quality of life instrument for children and adolescents. Gac. Sanit. 19, 93–102.
Beck, A., LeBlanc, J.C., Morissette, K., Hamel, C., Skidmore, B., Colquhoun, H., Lang, E., Moore, A., Riva, J.J., Thombs, B.D., Patten, S., Bragg, H., Colman, I., Goldfield, G.S., Nicholls, S.G., Pajer, K., Potter, B.K., Mendez, R., Vasa, F., Hutton, B., Shea, B.J., Graham, E., Little, J., Moher, D., Stevens, A., 2021. Screening for depression in children and adolescents: a protocol for a systematic review update. Syst. Rev. 10 https://doi.org/10.1186/s13643-020-01568-3.
Bernaras, E., Jaureguizar, J., Garaigordobil, M., 2019. Child and adolescent depression: a review of theories, evaluation instruments, prevention programs, and treatments. Front. Psychol. 10, 543. https://doi.org/10.3389/FPSYG.2019.00543.
Bianchi, R., Verkuilen, J., Toker, S., Schonfeld, I.S., Gerber, M., Brähler, E., Kroenke, K., 2022. Is the PHQ-9 a unidimensional measure of depression? A 58,272-participant study. Psychol. Assess. 34, 595–603. https://doi.org/10.1037/PAS0001124.
Boothroyd, L., Dagnan, D., Muncer, S., 2019. PHQ-9: one factor or two? Psychiatry Res. 271, 532–534. https://doi.org/10.1016/J.PSYCHRES.2018.12.048.
Borghero, F., Martínez, V., Zitko, P., Vöhringer, P.A., Cavada, G., Rojas, G., 2018. [Screening depressive episodes in adolescents. Validation of the Patient Health Questionnaire-9 (PHQ-9)]. Rev. Med. Chil. 146, 479–486. https://doi.org/10.4067/s0034-98872018000400479.
Burdzovic, A.J., Brunborg, G.S., 2017. Depressive symptomatology among Norwegian adolescent boys and girls: the patient health Questionnaire-9 (PHQ-9) psychometric properties and correlates. Front. Psychol. 8, 887. https://doi.org/10.3389/fpsyg.2017.00887.
Cheung, G.W., Rensvold, R.B., 2002. Evaluating goodness-of-fit indexes for testing measurement invariance. Struct. Equ. Model 9, 233–255. https://doi.org/10.1207/S15328007SEM0902_5.
Curri, K.E., Thompson, J., Dennis, A., 2022. Bridging the gap in adolescent depression: an action plan intervention. J. Pediatr. Health Care 36, 284–289. https://doi.org/10.1016/J.PEDHC.2021.10.006.
Davis, M., Jones, J.D., So, A., Benton, T.D., Boyd, R.C., Melhem, N., Ryan, N.D., Brent, D. A., Young, J.F., 2022. Adolescent depression screening in primary care: who is screened and who is at risk? J. Affect. Disord. 299, 318–325. https://doi.org/10.1016/J.JAD.2021.12.023.
El-Den, S., Chen, T.F., Gan, Y.L., Wong, E., O'Reilly, C.L., 2018. The psychometric properties of depression screening tools in primary healthcare settings: a systematic review. J. Affect. Disord. 225, 503–522. https://doi.org/10.1016/J.JAD.2017.08.060.
Eaton, N.R., Bringmann, L.F., Elmer, T., Fried, E.I., Forbes, M.K., Greene, A.L., Krueger, R.F., Kotov, R., McGorry, P.D., Mei, C., Waszczuk, M.A., 2023. A review of approaches and models in psychopathology conceptualization research. Nat. Rev. Psychol. 1–15. https://doi.org/10.1038/s44159-023-00218-4.
Elosua, P., Aguado, D., Fonseca-Pedrero, E., Abad, F.J., Santamaría, P., 2023. New trends in digital technology-based psychological and educational assessment. Psicothema 35, 50–57. https://doi.org/10.7334/PSICOTHEMA2022.241.
Fatiregun, A.A., Kumapayi, T.E., 2014. Prevalence and correlates of depressive symptoms among in-school adolescents in a rural district in southwest Nigeria. J. Adolesc. 37, 197–203. https://doi.org/10.1016/j.adolescence.2013.12.003.
Finning, K., Ukoumunne, O.C., Ford, T., Danielsson-Waters, E., Shaw, L., Romero De Jager, I., Stentiford, L., Moore, D.A., 2019. The association between child and adolescent depression and poor attendance at school: a systematic review and meta-analysis. J. Affect. Disord. 245, 928–938. https://doi.org/10.1016/J.JAD.2018.11.055.
Fonseca-Pedrero, E., Lemos-Giráldez, S., Paino, M., Villazón-García, U., Muñiz, J., 2009. Validation of the schizotypal personality questionnaire brief form in adolescents. Schizophr. Res. 111, 53–60. https://doi.org/10.1016/j.schres.2009.03.006.
Fonseca-Pedrero, E., Díez-Gómez, A., Pérez-Albéniz, A., Lucas-Molina, B., Al-Halabí, S., Calvo, P., 2023a. Psychology professionals in educational contexts: an unavoidable necessity. Papeles del Psicólogo/Psychologist Papers 44, 112–124. https://doi.org/10.23923/pap.psicol.3018.
Fonseca-Pedrero, E., Pérez-Albéniz, A., Al-Halabí, S., Lucas-Molina, B., Ortuño-Sierra, J., Díez-Gómez, A., Pérez-Sáenz, J., Inchausti, F., García, A.V.V., García, A.G., Solana, R.A., Ródenas-Perea, G., Clemente, M.P.D.V., López, A.C., Debbané, M., 2023b. PSICE Project Protocol: evaluation of the unified protocol for transdiagnostic treatment for adolescents with emotional symptoms in school settings. Clínica y Salud 34, 15–22. https://doi.org/10.5093/CLYSA2023A3.
Forman-Hoffman, V., McClure, E., McKeeman, J., Wood, C.T., Middleton, J.C., Skinner, A.C., Perrin, E.M., Viswanathan, M., 2016. Screening for major depressive disorder in children and adolescents: a systematic review for the U.S. preventive services task force. Ann. Intern. Med. 164, 342–349. https://doi.org/10.7326/M15-2259.
GBD 2019 Mental Disorders Collaborators, 2022. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet Psychiat. 9, 137–150. https://doi.org/10.1016/S2215-0366(21)00395-3.
González-Blanch, C., Medrano, L.A., Muñoz-Navarro, R., Ruíz-Rodríguez, P., Moriana, J. A., Limonero, J.T., Schmitz, F., Cano-Vindel, A., 2018. Factor structure and measurement invariance across various demographic groups and over time for the PHQ-9 in primary care patients in Spain. PLoS ONE 13, e0193356. https://doi.org/10.1371/JOURNAL.PONE.0193356.
González-Roz, A., Martínez-Loredo, V., Maalouf, W., Fernández-Hermida, J.R., Al-Halabí, S., 2023. Protocol for a trial assessing the efficacy of a universal school-based prevention program for addictive behaviors. Psicothema 35, 41–49. https://doi.org/10.7334/PSICOTHEMA2022.151.
Goodman, R., 1997. The strengths and difficulties questionnaire: a research note. J. Child Psychol. Psychiatry 38, 581–586.
Guo, X., Zhu, G.R., 2021. Universal depression screening in schools—promises and challenges in addressing adolescent mental health need. JAMA Netw. Open 4, e2132858. https://doi.org/10.1001/jamanetworkopen.2021.32858.
Hambleton, R., Swaminathan, H., Rogers, J., 1991. Fundamentals of Item Response Theory. Sage.
Hankin, B.L., 2015. Depression from childhood through adolescence: risk mechanisms across multiple systems and levels of analysis. Curr. Opin. Psychol. https://doi.org/10.1016/j.copsyc.2015.01.003.
Hankin, B.L., Fraley, R.C., Lahey, B.B., Waldman, I.D., 2005. Is depression best viewed as a continuum or discrete category? A taxometric analysis of childhood and adolescent depression in a population-based sample. J. Abnorm. Psychol. 114, 96–110. https://doi.org/10.1037/0021-843X.114.1.96.
Horn, J.L., McArdle, J.J., 1992. A practical and theoretical guide to measurement equivalence in aging research. Exp. Aging Res. 18, 117–144.
Hu, L.-T., Bentler, P.M., 1999. Cut off criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct. Equ. Model. 6, 1–55.
Johnson, J.G., Harris, E.S., Spitzer, R.L., Williams, J.B.W., 2002. The Patient Health Questionnaire for adolescents: validation of an instrument for the assessment of mental disorders among adolescent primary care patients. J. Adolesc. Heal. 30, 196–204. https://doi.org/10.1016/S1054-139X(01)00333-0.
Kotov, R., Krueger, R.F., Watson, D., Cicero, D.C., Conway, C.C., DeYoung, C.G., Eaton, N.R., Forbes, M.K., Hallquist, M.N., Latzman, R.D., Mullins-Sweatt, S.N.,

Ruggero, C.J., Simms, L.J., Waldman, I.D., Waszczuk, M.A., Wright, A.G.C., 2021. The hierarchical taxonomy of psychopathology (HiTOP): a quantitative nosology based on consensus of evidence. Annu. Rev. Clin. Psychol. 17, 83–108. https://doi.org/10.1146/annurev-clinpsy-081219-093304.
Kroenke, K., 2021. PHQ-9: global uptake of a depression scale. World Psychiatr. 20, 135–136. https://doi.org/10.1002/WPS.20821.
Kroenke, K., Spitzer, R.L., Williams, J.B.W., 2001. The PHQ-9: validity of a brief depression severity measure. J. Gen. Intern. Med. 16, 606–613. https://doi.org/10.1046/J.1525-1497.2001.016009606.X.
Leung, D.Y.P., Mak, Y.W., Leung, S.F., Chiang, V.C.L., Loke, A.Y., 2020. Measurement invariances of the PHQ-9 across gender and age groups in Chinese adolescents. Asia Pac. Psychiatry 12. https://doi.org/10.1111/APPY.12381.
Muñoz-Navarro, R., Cano-Vindel, A., Moriana, J.A., Medrano, L.A., Ruiz-Rodríguez, P., Agüero-Gento, L., Rodríguez-Enríquez, M., Pizà, M.R., Ramírez-Manent, J.I., 2017. Screening for generalized anxiety disorder in Spanish primary care centers with the GAD-7. Psychiatry Res. 256, 312–317. https://doi.org/10.1016/J.PSYCHRES.2017.06.023.
NICE Guidance, 2019. Depression in Children and Young people: Identification and Management. NICE guideline [NG134].
Olino, T.M., Yu, L., Klein, D.N., Rohde, P., Seeley, J.R., Pilkonis, P.A., Lewinsohn, P.M., 2012. Measuring depression using item response theory: an examination of three measures of depressive symptomatology. Int. J. Method. Psychiatr. Res. 21, 76–85. https://doi.org/10.1002/mpr.1348.
Ortuño-Sierra, J., Sebastián-Enesco, C., Pérez-Albéniz, A., Lucas-Molina, B., Fonseca-Pedrero, E., 2022. Spanish normative data of the Strengths and Difficulties Questionnaire in a community-based sample of adolescents. Int. J. Clin. Health Psychol. 22 https://doi.org/10.1016/J.IJCHP.2022.100328.
Patra, A.P., Kumar, K., 2023. Screening for depression and suicide in children. StatPearls.
Piqueras, J.A., Garcia-Olcina, M., Rivera-Riquelme, M., Martinez-Gonzalez, A.E., Cuijpers, P., 2022. DetectaWeb-distress scale: a global and multidimensional web-based screener for emotional disorder symptoms in children and adolescents. Front. psychol. 12, 627604 https://doi.org/10.3389/fpsyg.2021.627604.
Ravens-Sieberer, U., Erhart, M., Rajmil, L., Herdman, M., Auquier, P., Bruil, J., Power, M., Duer, W., Abel, T., Czemy, L., Mazur, J., Czimbalmos, A., Tountas, Y., Hagquist, C., Kilroe, J., 2010. Reliability, construct and criterion validity of the KIDSCREEN-10 score: a short measure for children and adolescents' well-being and health-related quality of life. Qual. Life Res. 19, 1487–1500. https://doi.org/10.1007/S11136-010-9706-5.
Rhew, I.C., Monahan, K.C., Oesterle, S., Hawkins, J.D., 2016. The communities that care brief depression scale: psychometric properties and criterion validity. J. Commun. Psychol. 44, 391–398. https://doi.org/10.1002/jcop.21766.
Richardson, L.P., McCauley, E., Grossman, D.C., McCarty, C.A., Richards, J., Russo, J.E., Rockhill, C., Katon, W., 2010. Evaluation of the Patient Health Questionnaire (PHQ-9) for detecting major depression among adolescents. Pediatrics 126, 1117. https://doi.org/10.1542/PEDS.2010-0852.
Riglin, L., Petrides, K.V., Frederickson, N., Rice, F., 2014. The relationship between emotional problems and subsequent school attainment: a meta-analysis. J. Adolesc. 37, 335–346. https://doi.org/10.1016/J.ADOLESCENCE.2014.02.010.
Shorey, S., Ng, E.D., Wong, C.H.J., 2022. Global prevalence of depression and elevated depressive symptoms among adolescents: a systematic review and meta-analysis. Br. J. Clin. Psychol. 61, 287–305. https://doi.org/10.1111/BJC.12333.
Sinclair-McBride, K., Morelli, N., Gusman, M., 2018. PHQ-9 administration in outpatient adolescent psychiatry services. Psychiatr. Serv. 69, 837–838. https://doi.org/10.1176/appi.ps.201800245.
Solmi, M., Radua, J., Olivola, M., Croce, E., Soardo, L., Salazar de Pablo, G., Il Shin, J., Kirkbride, J.B., Jones, P., Kim, J.H., Kim, J.Y., Carvalho, A.F., Seeman, M.V., Correll, C.U., Fusar-Poli, P., 2022. Age at onset of mental disorders worldwide: large-scale meta-analysis of 192 epidemiological studies. Mol. Psychiatry 27, 281–295. https://doi.org/10.1038/S41380-021-01161-7.
Spitzer, R.L., Kroenke, K., Williams, J.B.W., 1999. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary care evaluation of mental disorders. Patient Health Questionnaire. JAMA 282, 1737–1744. https://doi.org/10.1001/JAMA.282.18.1737.
Spitzer, R.L., Kroenke, K., Williams, J.B.W., Löwe, B., 2006. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch. Intern. Med. 166, 1092–1097. https://doi.org/10.1001/ARCHINTE.166.10.1092.
Thapar, A., Eyre, O., Patel, V., Brent, D., 2022. Depression in young people. Lancet 400, 617–631. https://doi.org/10.1016/S0140-6736(22)01012-1.
Tsai, F.J., Huang, Y.H., Liu, H.C., Huang, K.Y., Huang, Y.H., Liu, S.I., 2014. Patient health questionnaire for school-based depression screening among Chinese adolescents. Pediatrics 133, e402–e409. https://doi.org/10.1542/peds.2013-0204.
WHO, 2021. WHO Guideline On School Health Services. WHO.
World Health Organization, 2022. World mental health report: transforming mental health for all. WHO.

ELSEVIER

Journal of Adolescent Health 39 (2006) 244–251

JOURNAL OF
ADOLESCENT
HEALTH

Original article

Does Body Satisfaction Matter? Five-year Longitudinal Associations between Body Satisfaction and Health Behaviors in Adolescent Females and Males

Dianne Neumark-Sztainer, Ph.D.[a,*], Susan J. Paxton, Ph.D.[b], Peter J. Hannan, M.Stat., Jess Haines, M.H.Sc., and Mary Story, Ph.D.

[a]Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, Minnesota
[b]School of Psychological Science, La Trobe University, Melbourne, Australia

Manuscript received September 15, 2005; manuscript accepted November 8, 2005

Abstract: **Purpose:** This study addresses the question, "Does body satisfaction matter?" by examining longitudinal associations between body satisfaction and weight-related health-promoting and health-compromising behaviors five years later among adolescents.

Methods: Project EAT-II followed an ethnically and socioeconomically diverse sample of 2516 adolescents from 1999 (Time 1) to 2004 (Time 2). Associations between body satisfaction at Time 1 and health behaviors at Time 2 were examined, adjusting for sociodemographic characteristics and Time 1 health behaviors, with and without adjustment for [body mass index] (BMI).

Results: In females, lower body satisfactio[n] unhealthy weight control behaviors and bing[e] and vegetable intake. After adjusting for BM[I] very unhealthy weight control behaviors, and males, lower body satisfaction predicted hig[her] unhealthy weight control behaviors, binge e[ating] activity. After adjusting for BMI, association[s] weight control behavior, and binge eating rema[in]

Conclusions: The study findings indicate th[at] serve as a motivator for engaging in healthy w[eight] the use of behaviors that may place adolescents Interventions with adolescents should strive to likely to lead to decreases in body satisfaction. rights reserved.

Sxls

Keywords: Body satisfaction; Body image; Adolescence; Obesity; [...] Dietary intake

High percentages of adolescents, particularly adolescent females, are dissatisfied with their bodies [1,2]. The high [...] dissatisfaction during adolescence, a [cr]itical period of identity formation, is disturbing in that body image, self-image, and self-esteem tend to be closely intertwined [3]. Longitudinal analyses show that low body satisfaction during early and middle adolescence is predictive of later signs of more global mental distress, including lower self-esteem and depressive symptoms [4–6]. Body dissatisfaction and preoccupation with thinness are strong

*Address correspondence to: Dr. Dianne Neumark-Sztainer, Division of Epidemiology and Community Health School of Public Health, University of Minnesota, 1300 South Second Street, Suite 300, Minneapolis, MN 55454.
E-mail address: neumark@epi.umn.edu

1054-139X/06/$ – see front matter
doi:10.1016/j.jadohealth.2005.12.001

ELSEVIER

Journal of Adolescent Health 39 (2006) 244–251

JOURNAL OF ADOLESCENT HEALTH

Original article

Does Body Satisfaction Matter? Five-year Longitudinal Associations between Body Satisfaction and Health Behaviors in Adolescent Females and Males

Dianne Neumark-Sztainer, Ph.D.[a,*], Susan J. Paxton, Ph.D.[b], Peter J. Hannan, M.Stat., Jess Haines, M.H.Sc., and Mary Story, Ph.D.

[a]*Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, Minnesota*
[b]*School of Psychological Science, La Trobe University, Melbourne, Australia*

Manuscript received September 15, 2005; manuscript accepted November 8, 2005

Abstract: **Purpose:** This study addresses the question, "Does body satisfaction matter?" by examining longitudinal associations between body satisfaction and weight-related health-promoting and health-compromising behaviors five years later among adolescents.

Methods: Project EAT-II followed an ethnically and socioeconomically diverse sample of 2516 adolescents from 1999 (Time 1) to 2004 (Time 2). Associations between body satisfaction at Time 1 and health behaviors at Time 2 were examined, adjusting for sociodemographic characteristics and Time 1 health behaviors, with and without adjustment for body mass index (BMI).

Results: In females, lower body satisfaction predicted higher levels of dieting, unhealthy and very unhealthy weight control behaviors and binge eating, and lower levels of physical activity and fruit and vegetable intake. After adjusting for BMI, associations between body satisfaction and dieting, very unhealthy weight control behaviors, and physical activity remained statistically significant. In males, lower body satisfaction predicted higher levels of dieting, healthy, unhealthy, and very unhealthy weight control behaviors, binge eating, and smoking, and lower levels of physical activity. After adjusting for BMI, associations between body satisfaction and dieting, unhealthy weight control behavior, and binge eating remained statistically significant.

Conclusions: The study findings indicate that, in general, lower body satisfaction does not serve as a motivator for engaging in healthy weight management behaviors, but rather predicts the use of behaviors that may place adolescents at risk for weight gain and poorer overall health. Interventions with adolescents should strive to enhance body satisfaction and avoid messages likely to lead to decreases in body satisfaction.

Keywords: Body satisfaction; Body image; Adolescence; Obesity; Eating disorders; Dieting; Smoking; Physical activity; Dietary intake

High percentages of adolescents, particularly adolescent females, are dissatisfied with their bodies [1,2]. The high prevalence of body dissatisfaction during adolescence, a critical period of identity formation, is disturbing in that body image, self-image, and self-esteem tend to be closely intertwined [3]. Longitudinal analyses show that low body satisfaction during early and middle adolescence is predictive of later signs of more global mental distress, including lower self-esteem and depressive symptoms [4–6]. Body dissatisfaction and preoccupation with thinness are strong

*Address correspondence to: Dr. Dianne Neumark-Sztainer, Division of Epidemiology and Community Health School of Public Health, University of Minnesota, 1300 South Second Street, Suite 300, Minneapolis, MN 55454.

E-mail address: neumark@epi.umn.edu

1054-139X/06/$ – see front matter © 2006 Society for Adolescent Medicine. All rights reserved.
doi:10.1016/j.jadohealth.2005.12.001

predictors of eating disorders and related disordered eating behaviors [7–9]. Thus, there is concern among adolescent health professionals about the high prevalence of adolescents with low levels of body satisfaction [10].

Given the widespread prevalence of obesity in adolescents [11] and its associated health consequences [12], an important question relates to the impact of body satisfaction on behaviors with implications for weight management [13,14]. Specifically, are adolescents with low levels of body satisfaction less likely to engage in healthy weight management behaviors, such as increased fruit and vegetable intake or regular physical activity, than adolescents who feel good about their bodies? Are they at increased risk for dieting, unhealthy weight control behaviors, or binge eating, which are associated with weight gain over time [15–18]?

Alternatively, are there advantages to not being satisfied with one's body? Can body dissatisfaction serve as a self-motivator to engage in healthier eating and physical activity behaviors? Heinberg and colleagues have argued that some level of body image dissatisfaction may be beneficial for individuals with average or above-average body mass index (BMI) values, because it may lead to healthy weight management behaviors [13,19]. They assert that the relationship between body image dissatisfaction and healthy weight management behaviors may be illustrated by an inverted U-shaped curve. When body image distress is very low, individuals may not engage in healthy eating and exercise behaviors, even if necessary to improve health outcomes. When body image distress is very high, individuals may fail to engage in healthy weight management behaviors because of a perceived inability to make meaningful changes in their bodies, or may engage in unhealthy dieting behaviors in a desperate attempt to lose weight.

The question as to whether body satisfaction predicts health-promoting or health-compromising behaviors has important implications for the design of interventions aimed at obesity prevention and overall health promotion among adolescents. The current study expands upon the growing body of literature exploring associations between body satisfaction and different health behaviors to address the research question, "Does body satisfaction matter, and if so, how?" We examine longitudinal associations between body image and a range of weight-related health-promoting and health-compromising behaviors five years later in an ethnically and socioeconomically diverse population of adolescents. We examine whether associations between body satisfaction and behavioral outcomes are approximately linear or have a U-shaped association in order to address the question raised by Heinberg and colleagues [13,19]. Analyses are done with and without adjustment for BMI, to determine whether associations between body satisfaction and outcome behaviors are a function of differences in weight status, and whether they also exist independently of BMI.

Methods

Study design and population

Project EAT-II is a longitudinal, follow-up study of Project EAT-I, a study of the socio-environmental, personal, and behavioral determinants of dietary intake and weight status in adolescents [1,20–22]. In Project EAT-I, 4746 junior and senior high school students in 31 Minnesota schools completed in-class surveys and anthropometric measures during the 1998–1999 academic year. Project EAT-II aimed to re-survey all original participants by mail to examine changes in their eating patterns and weight status five years later (2003–2004). The University of Minnesota's Institutional Review Board Human Subjects Committee approved all study protocols.

Of the original study population, 1074 (22.6%) were lost to follow-up for various reasons, primarily missing contact information at EAT-I (n = 411) and no address found at follow-up (n = 591). Of the remaining 3672 participants contacted by mail, 2516 completed surveys, representing 53.0% of the original cohort and 68.4% of participants who could be contacted for Project EAT-II. The final study population consisted of 1130 males (44.9%) and 1386 females (55.1%) who completed surveys for both EAT-I (Time 1) and EAT-II (Time 2). One-third of the participants (32.0%) were in the younger cohort; at Time 1 their mean age was 12.8 years (SD = .8) and at Time 2 their mean age was 17.2 years (SD = .6). Two-thirds of the participants (68.0%) were in the older cohort; at Time 1 their mean age was 15.8 years (SD = .8) and at Time 2 their mean age was 20.4 years (SD = .8).

Measures

Body satisfaction. Body satisfaction was assessed with a modified version of the Body Shape Satisfaction Scale [23]. Ten items assessed satisfaction with different body parts (height, weight, body shape, waist, hips, thighs, stomach, face, body build, shoulders). For each item there were five Likert response categories ranging from "very dissatisfied" (1) to "very satisfied" (5) (Cronbach alpha = .92). Responses were summed with higher scores indicative of higher levels of body satisfaction. Participants were divided into quartiles (low, low-middle, high-middle, high) based upon distribution of scores in the total EAT-II sample (gender-combined).

Dieting and weight control behaviors. Dieting was assessed with the question: "How often have you gone on a diet during the last year? By "diet" we mean changing the way you eat so you can lose weight" [1]. Responses were dichotomized: no (never) and yes (any frequency). Specific types of weight control behaviors were assessed with the question: "Have you done any of the following things in order to lose weight or keep from gaining weight during the past year? (yes or no for each method)" [1]. *Healthy weight*

control behaviors included: exercised, ate more fruits and vegetables, ate less high-fat foods, and ate less sweets. *Unhealthy weight control behaviors* included: fasted, ate very little food, used a food substitute (powder or a special drink), skipped meals, and smoked more cigarettes. *Very unhealthy weight control behaviors* included: took diet pills, made myself vomit, used laxatives, and used diuretics.

Binge eating. Binge eating was assessed with the question: "In the past year, have you ever eaten so much food in a short period of time that you would be embarrassed if others saw you (binge eating)?" (yes/no) [1].

Smoking. Participants were asked about the frequency with which they smoked cigarettes over the past year. Response categories included: never, a few times, monthly, weekly, and daily. Responses were dichotomized for analysis; adolescents who reported that they never smoked or had smoked only a few times were compared with adolescents who reported smoking at least monthly.

Physical activity. Moderate-to-vigorous physical activity (MVPA) was assessed with a modified version of the Leisure Time Exercise Questionnaire [24]. Two questions were asked to assess how many hours were spent in strenuous (e.g., biking fast, aerobic dancing, or running) or moderate (e.g., walking quickly, baseball, or gymnastics) physical activity behaviors in a usual week. The responses (0, <.5, .5–2, 2.5–4, 4.5–6, and >6 hours/week) were recoded for analyses (0, .3, 1.3, 3.3, 5.3, and 8 hours/week).

Fruit and vegetable intake. Fruit and vegetable intake was assessed with the Youth and Adolescent Food Frequency Questionnaire (YAQ). The YAQ has been tested for reproducibility and has been compared with averages from three 24-hour dietary recalls, and findings have been within acceptable ranges for dietary assessment tools [25,26]. Mean daily servings of fruits and vegetables were calculated from responses to questions assessing frequencies of intakes of specific fruits and vegetables. French fries were excluded from total servings.

Body mass index (BMI). At Time 1, height and weight measurements were taken by trained research staff in a private area within the participants' schools, using standardized equipment and procedures [27]. BMI was derived from the formula: weight in kilograms divided by the square of height in meters.

Sociodemographic characteristics. Gender, age, ethnicity/race, and *socioeconomic status (SES)* were based on self-report at Time 1. The prime determinant of SES was parental educational level, defined by the higher level of educational attainment of either parent. An algorithm was developed that also took into account family eligibility for public assistance, eligibility for free or reduced-cost school meals, and employment status of the mother and father [21].

Age cohort was based on grade in school (middle school vs. high school) at Time 1.

Statistical analysis

Descriptive summaries (percentages and means) for body satisfaction at Time 1 and health behaviors at Time 2 were first examined. We treated two behavioral outcomes (physical activity and fruit and vegetable intake) as continuous measures, whereas the other behavioral outcomes were dichotomized. Within gender, four groups were then identified on the basis of Time 1 body satisfaction quartiles: low, low-middle, high-middle and high body satisfaction. A series of general linear models was conducted in which each health behavior was the outcome variable, body satisfaction group was the independent variable, and age-cohort, race, SES, age-in-years (to accommodate the unequal follow-up times), and Time 1 of the relevant health behavior were entered as covariates. To examine whether differences across body satisfaction groups in Time 2 health behaviors were related to BMI, a second set of analyses was conducted in which BMI was also entered as a covariate. We obtained adjusted mean behavioral outcomes at Time 2 in the four categories of body satisfaction from each model. For the dichotomous outcomes, the adjusted means are direct estimates of the adjusted probabilities of the Time 2 behavior by quartile of body satisfaction. We tested a priori for linear trend across the quartiles. We also examined behavioral outcomes by the four levels of body satisfaction for any nonlinear patterns and tested for quadratic relationships; for this secondary hypothesis, p values are provided only when there was evidence of a U-shaped association.

Attrition in the study population was not equal across sociodemographic characteristics. Time 2 participants were more likely to be female, white, and of higher SES than Time 1 participants. Thus, in all analyses, the data were weighted to adjust for differential response rates using the response propensity method [28] by which the inverse of the estimated probability that an individual responded at Time 2 was used as the weight. The weighting method results in estimates representative of the demographic makeup of the original Project EAT-I sample. The weighted ethnic, racial and SES proportions of the study population are as follows: 48.3% white, 18.9% African American, 5.8% Hispanic, 19.6% Asian, 3.6% Native American, and 3.8% mixed or other race, whereas SES was low (17.8%), middle-low (18.9%), middle (26.7%), middle-high (23.3%), and high (13.3%). After weighting, we compared responders to the Project EAT-II survey with nonresponders for the variables being examined in the current analysis. In girls, responders had higher MVPA than nonresponders (5.7 vs. 5.0 h/week, respectively) and in boys, responders reported lower levels of very unhealthy weight control behaviors than nonresponders (5.1% vs. 9.2%, respectively); no response bias

Table 1
Time 1 body satisfaction and Time 2 health behaviors: percentages and mean values in females and males

	Females n = 1341–1377 % or mean (SD)	Males n = 1095–1121 % or mean (SD)
Body satisfaction (Mean score range: 10–50)	31.76 (9.55)	36.93 (8.73)
Dieting (% past year)	56.3%	26.7%
Healthy weight control (% past year)	85.3%	63.1%
Unhealthy weight control (% past year)	62.7%	33.1%
Very unhealthy weight control (% past year)	21.9%	6.7%
Binge eating (% past year)	15.5%	5.1%
Smoking (% monthly)	28.4%	29.3%
Physical activity (MVPA) (Mean hours/week)	3.93 (3.56)	6.11 (4.27)
Fruit and vegetable intake (Mean servings/day)	3.59 (2.48)	3.13 (2.32)

was found for body satisfaction or for any of the other outcomes in either gender.

Results

Prevalence of Time 1 body satisfaction and Time 2 health behaviors

Mean body satisfaction scores at Time 1 are shown in Table 1. Among females, the distribution of body satisfaction based upon quartiles in the total EAT-II sample was as follows: low (33.8%, n = 444), low-middle (25.6%, n = 335), high-middle (22.2%, n = 291), and high (18.4%, n = 241). Among males, the distribution by quartiles was as follows: low (23.7%, n = 312), low-middle (25.9%, n = 340), high-middle (24.1%, n = 316), and high (26.3%, n = 344). At Time 2, high percentages of respondents, particularly females, reported dieting and weight control behaviors. Three times more females than males reported binge eating. Approximately one-quarter of the females and males reported monthly smoking. Females reported about four hours of MVPA a week, whereas males reported about six hours a week. Mean intakes of fruits and vegetables were 3.6 servings a day in females and 3.1 servings a day in males.

Females: associations between Time 1 body satisfaction and Time 2 behaviors

In analyses examining associations between Time 1 body satisfaction and Time 2 health behaviors, adjusted for Time 1 behaviors and race, SES, and age, body satisfaction predicted higher levels of dieting, unhealthy weight control behaviors, very unhealthy weight control behaviors and binge eating, and lower levels of MVPA and fruit/vegetable intake (Table 2). In analyses that also adjusted for BMI, associations remained statistically significant for dieting, unhealthy weight control behaviors, and MVPA. Thus, lower body satisfaction predicted higher levels of dieting and unhealthy weight control behaviors and lower levels of MVPA, independently of BMI (Table 2). An examination of behavioral outcomes by the four levels of body satisfaction for nonlinear patterns revealed U-shaped curves only for healthy weight control behaviors. Among girls with the lowest and highest levels of body satisfaction, 89% and 87%, respectively, reported using healthy weight control behaviors, versus 83% of the girls in the mid-categories of body satisfaction (p value for quadratic association = .008);

Table 2
Females: Time 2 health behaviors by Time 1 body satisfaction: adjusted percentages and means

	n	Dieting % past year	Healthy weight control % past year	Unhealthy weight control % past year	Very unhealthy weight control % past year	Binge eating % past year	Smoking % monthly	Physical activity (MVPA) Hours/wk M (SD)	Fruit and vegetable intake Servings/day M (SD)
Body satisfaction[a]									
Low	386–433	65.1	88.8	68.2	32.7	19.8	30.5	3.92 (.16)	3.35 (.10)
Low-middle	292–330	58.4	82.7	66.4	20.0	14.5	29.0	3.78 (.18)	3.36 (.11)
High-middle	253–279	52.6	83.1	57.5	15.7	12.4	27.1	4.22 (.20)	3.51 (.12)
High	201–227	47.1	87.2	53.3	15.7	12.3	26.5	4.42 (.22)	3.69 (.14)
p Value		< .001	.628	< .001	< .001	.012	.190	.028	.037
Body satisfaction[b]									
Low	347–383	59.0	85.8	62.0	27.5	17.5	29.4	3.92 (.18)	3.40 (.11)
Low-middle	273–300	57.4	81.6	65.6	19.9	13.3	28.4	3.79 (.19)	3.31 (.12)
High-middle	239–257	54.2	83.8	60.1	17.2	13.7	27.8	4.25 (.21)	3.49 (.13)
High	191–211	50.4	88.4	57.4	19.8	14.3	25.9	4.45 (.23)	3.69 (.15)
p Value		.041	.317	.164	.029	.385	.338	.034	.079

[a] Adjusted for race, SES, age, and Time 1 behaviors.
[b] Adjusted for race, SES, age, Time 1 behaviors, and BMI.

Table 3
Males: Time 2 health behaviors by Time 1 body satisfaction: adjusted percentages and means

	n	Dieting % past year	Healthy weight control % past year	Unhealthy weight control % past year	Very unhealthy weight control % past year	Binge eating % past year	Smoking % monthly	Physical activity (MVPA) Hours/wk M (SD)	Fruit and vegetable intake Servings/day M (SD)
Body satisfaction[a]									
Low	116–130	46.2	74.6	50.1	13.9	11.3	40.0	5.72 (.36)	3.32 (.20)
Low-middle	231–262	30.9	64.7	34.0	6.3	4.8	30.5	6.12 (.25)	2.99 (.14)
High-middle	258–293	24.8	62.6	33.6	5.7	3.8	29.9	6.25 (.23)	3.12 (.13)
High	307–347	18.4	55.5	24.0	4.7	3.1	28.4	6.67 (.22)	3.13 (.12)
p-value		< .001	< .001	< .001	< .001	< .001	.011	< .024	.535
Body satisfaction[b]									
Low	106–115	36.4	64.4	40.1	7.1	11.0	35.8	6.15 (.39)	3.44 (.21)
Low-middle	219–238	29.4	64.0	32.1	5.8	3.3	27.0	5.95 (.26)	2.86 (.14)
High-middle	243–270	23.3	62.7	33.2	5.8	3.3	29.6	6.17 (.24)	3.09 (.13)
High	288–324	21.1	58.2	26.7	6.0	3.4	29.2	6.51 (.23)	3.09 (.12)
p Value		< .001	.187	.009	.690	.002	.244	.376	.060

[a] Adjusted for race, SES, age, and Time 1 behaviors.
[b] Adjusted for race, SES, age, Time 1 behaviors, and BMI.

after adjusting for BMI the pattern persisted (p value for quadratic association = .031).

Males: associations between Time 1 body satisfaction and Time 2 behaviors

In analyses examining associations between Time 1 body satisfaction and Time 2 health behaviors, adjusting for Time 1 behaviors and demographics, lower body satisfaction predicted higher levels of dieting, healthy, unhealthy, and very unhealthy weight control behaviors, binge eating, and smoking and lower levels of MVPA (Table 3). After adjusting for BMI, associations between body satisfaction and dieting, unhealthy weight control behaviors, and binge eating remained statistically significant (Table 3). An examination of behavioral outcomes by the four levels of body satisfaction for nonlinear patterns did not reveal any U-shaped patterns of association.

Discussion

The current study examined five-year longitudinal associations between body satisfaction and an array of health-related behaviors among adolescents in order to address the question, "Does body satisfaction matter?" Our findings indicate that lower levels of body satisfaction are associated with more health-compromising behaviors, such as unhealthy weight control behaviors and binge eating, and fewer health-promoting behaviors, such as physical activity. Having a lower level of body satisfaction did not incur advantages in terms of behavioral outcomes, with the exception of reported healthy weight control behaviors, which occurred in conjunction with unhealthy weight control behaviors. Thus, the picture that emerges suggests that body satisfaction does matter and we need to be concerned about the high prevalence of adolescents who express body dissatisfaction.

Longitudinal analyses, adjusting for Time 1 behaviors and sociodemographic characteristics, but not adjusted for BMI, revealed strong and consistent patterns between body satisfaction and health-related behaviors five years later. In females, lower body satisfaction predicted higher levels of dieting, unhealthy and very unhealthy weight control behaviors, and binge eating, and lower levels of physical activity and fruit and vegetable intake. The only U-shaped association found among the females was between body satisfaction and healthy weight control behaviors. In males, lower body satisfaction predicted higher levels of dieting, unhealthy and very unhealthy weight control behaviors, binge eating, and smoking, and lower levels of physical activity. There was only one suggestion of a positive impact of low body satisfaction among males; body satisfaction was inversely associated with the reported use of healthy weight control behaviors five years later.

In similar analyses that also adjusted for BMI, a number of associations remained statistically significant. Independent of BMI, among females, lower levels of body satisfaction predicted higher levels of dieting and unhealthy weight control behaviors and lower levels of physical activity. Among males, low levels of body satisfaction predicted higher levels of dieting, unhealthy weight control behaviors, and binge eating. However, patterns of association tended to be weaker and less consistent in these analyses than in analyses that did not adjust for BMI, suggesting that, in part, associations between body satisfaction and health behaviors are a function of BMI.

As in the current study, other longitudinal studies have also shown that lower body satisfaction is associated with higher levels of dieting, dietary restraint, unhealthy weight

control behaviors and binge eating [5,8,18,29–31]. These relationships are disturbing, given findings from longitudinal studies showing that dieting, unhealthy weight control behaviors, and binge eating predict weight gain in adolescents, even after adjustment for baseline differences in weight status [15–18]. In light of the high prevalence of obesity, an important question is how can we help adolescents appreciate their bodies yet recognize the importance of striving for a healthy weight through healthier weight management behaviors?

Associations between body satisfaction and smoking tended to be weaker in the current study than in other cross-sectional [32–34] and longitudinal studies [35], particularly among females. One reason may be that most other studies have measured weight-specific attitudes or behaviors such as drive for thinness or dietary restraint, and not satisfaction with different parts of the body, as was done in the current analysis. Additional contributing factors include the long period between Time 1 assessment of body satisfaction and Time 2 smoking, and age and developmental factors whereby girls who were dissatisfied with their bodies may have already initiated smoking at Time 1. Indeed, in analyses unadjusted for Time 1 behaviors, lower levels of body satisfaction at Time 1 were significantly associated with higher levels of Time 2 smoking in females (data not shown). Our findings should not be interpreted as meaning that adolescents do not smoke for weight control purposes. As previously reported [1], nearly 5% of males and 10% of females in the Project EAT population indicated that they smoked more cigarettes at Time 1 to lose weight or avoid gaining weight.

Fewer studies have examined associations between body satisfaction and healthier weight management strategies, such as eating more fruits and vegetables and increasing physical activity. Longitudinal studies examining associations between body satisfaction and physical activity have not found statistically significant associations; however, these studies examined the independent contribution of body satisfaction, after adjusting for stronger predictor variables such as perceived athletic competence and parental support for physical activity [29,36,37]. Findings from the current study were consistent with other studies in finding that low body satisfaction is a not a motivator toward behaviors likely to be effective in long-term weight management. However, because of the potential complexity of these associations, further exploration would be informative. In the current study, we found different patterns of association between body satisfaction and the reported use of healthy behaviors during the past year such as increasing physical activity and fruit and vegetable intake for weight control reasons, than between body satisfaction and these behaviors as assessed using the leisure time exercise questionnaire and the food frequency questionnaire, respectively. In a cross-sectional study of college students, body dissatisfaction was associated with exercise for weight, tone, and attractiveness reasons, but not with exercising for mood, health, and enjoyment [38]. This raises the question as to how one's reasons for engaging in behaviors such as physical activity or fruit and vegetable intake influence the intensity, consistency, and duration with which behaviors are implemented.

Strengths of the current study include the large and diverse study population in terms of ethnicity and SES; the five-year follow-up during key transitional periods of adolescence; the assessment of BMI using measured values of height and weight; the assessment of body satisfaction using a psychometrically sound tool; and a broader assessment of health-related variables, particularly regarding weight control behaviors and dietary intake, than is typically done in large population-based surveys of youth. We are unaware of any other studies that have examined longitudinal associations between body satisfaction and a broad array of health behaviors in such a large and diverse population of adolescents as was done in the current study.

Although these strengths contribute to the uniqueness and utility of the findings, study limitations also need to be taken into account in interpreting the findings. First, in spite of multiple attempts to reach the original study participants, there was study attrition and participants in EAT-II differed from the original cohort. In order to enhance our ability to make extrapolations to a sample similar to the original study population, nonresponse weighting procedures were used. Second, given the comprehensive nature of our survey, some of the measures of health behaviors were brief and based on self-report on a survey, rather than on actual measurements, observations, or clinical assessment. Finally, although the longitudinal nature of the study allowed us to examine whether body satisfaction predicts health behaviors over time, we cannot establish causality. Interventions aimed at improving body satisfaction should include an assessment of health behaviors in order to determine whether change in body satisfaction leads to change in behavioral patterns. Future research should expand upon the current analysis by exploring associations between body satisfaction and behavioral outcomes within subgroups of the population, e.g., across weight status and ethnicity. It is also of interest to explore the impact of different aspects of body satisfaction (e.g., shoulder width, weight) within different subgroups. For example, dissatisfaction with shoulder width is likely to have a different impact on underweight boys than on overweight girls.

Findings from the present study clearly indicate the importance of body satisfaction for the overall well-being of adolescents. Body satisfaction was predictive of health-related behaviors even after a five-year period. Our findings suggest the importance of avoiding messages or interventions that may, inadvertently, lead to lower levels of body satisfaction in adolescents. Parents, educators, and health care providers should resist utilizing messages aimed at motivating adolescents toward behavioral change via de-

creasing their comfort with their bodies. Instead, it may be more effective to encourage positive change via messages that enhance body satisfaction and a desire to care for one's body. Recommendations regarding the types of messages that might be most effective cannot be made from the current analysis; rather, this research provides a justification for the exploration of suitable messages and an assessment of their effectiveness in both enhancing body satisfaction and the adoption of a healthier lifestyle. Programs aimed at promoting both a healthy weight and a positive body image may offer the most promise for decreasing potentially harmful behaviors, such as unhealthy weight control and binge eating, and in promoting the use of healthier behaviors, such as increased physical activity, which are more likely to be effective for long-term weight management and overall health promotion.

Acknowledgment

This study was supported by Grant R40 MC 00319 from the Maternal and Child Health Bureau (Title V, Social Security Act), Health Resources and Services Administration, Department of Health and Human Services.

References

[1] Neumark-Sztainer D, Croll J, Story M, et al. Ethnic/racial differences in weight-related concerns and behaviors among adolescent girls and boys: findings from Project EAT. J Psychosom Res 2002;53:963–74.
[2] Ricciardelli LA, McCabe MP. Dietary restraint and negative affect as mediators of body dissatisfaction and bulimic behavior in adolescent girls and boys. Behav Res Ther 2001;39:1317–28.
[3] Keery H, van den Berg P, Thompson JK. An evaluation of the tripartite influence model of body dissatisfaction and eating disturbance with adolescent girls. Body Image 2004;1:237–51.
[4] Holsen I, Kraft P, Roysamb E. The relationship between body image and depressed mood in adolescence: a 5-year longitudinal panel study. J Health Psychol 2001;6:613–27.
[5] Johnson F, Wardle J. Dietary restraint, body dissatisfaction, and psychological distress: a prospective analysis. J Abnorm Psychol 2005;114:119–25.
[6] Stice E, Bearman SK. Body-image and eating disturbances prospectively predict increases in depressive symptoms in adolescent girls: a growth curve analysis. Dev Psychol 2001;37:597–607.
[7] Cooley E, Toray T. Disordered eating in college freshman women: a prospective study. J Am Coll Health 1996;44:229–35.
[8] Stice E. A prospective test of the dual-pathway model of bulimic pathology: mediating effects of dieting and negative affect. J Abnorm Psychol 2001;110:124–35.
[9] The McKnight Investigators. Risk factors for the onset of eating disorders in adolescent girls: results of the McKnight longitudinal risk factor study. Am J Psychiatry 2003;160:248–54.
[10] Smolak L. Body image in children and adolescents: where do we go from here? Body Image 2004;1:15–28.
[11] Hedley AA, Ogden CL, Johnson CL, et al. Prevalence of overweight and obesity among US children, adolescents, and adults, 1999–2002. JAMA 2004;291:2847–50.
[12] U.S. Department of Health and Human Services. The Surgeon General's Call to Action to Prevent and Decrease Overweight and Obesity. Rockville, MD: Office of Disease Prevention and Health Promotion, Centers for Disease Control and Prevention, National Institutes of Health, 2001.
[13] Heinberg LJ, Thompson JK, Matzon JL. Body image dissatisfaction as a motivator for healthy lifestyle change: is some distress beneficial? In: Striegel-Moore RH, Smolak L, eds. Eating Disorders Innovative Directions in Research and Practice. Washington, DC: American Psychological Association, 2001:215–32.
[14] Schwartz MB, Brownell KD. Obesity and body image. Body Image 2004;1:43–56.
[15] Neumark-Sztainer D, Wall M, Guo J, et al. Obesity, disordered eating, and eating disorders in a longitudinal study of adolescents: how do dieters fare five years later? J Am Diet Assoc In Press.
[16] Field AE, Austin SB, Taylor CB, et al. Relation between dieting and weight change among preadolescents and adolescents. Pediatrics 2003;112:900–6.
[17] Stice E, Cameron RP, Killen JD, et al. Naturalistic weight-reduction efforts prospectively predict growth in relative weight and onset of obesity among female adolescents. J Consult Clin Psychol 1999;67:967–74.
[18] Stice E, Presnell K, Spangler D. Risk factors for binge eating onset in adolescent girls: a 2-year prospective investigation. Health Psychol 2002;21:131–8.
[19] Heinberg L. The Role of Body Image Distress in Obesity: Is it Motivating or Malevolent? Academy for Eating Disorders International Conference on Eating Disorders. Montreal, Quebec, Canada: Academy for Eating Disorders, 2005.
[20] Neumark-Sztainer D, Story M, Perry C, et al. Factors influencing food choices of adolescents: findings from focus-group discussions with adolescents. J Am Diet Assoc 1999;99:929–37.
[21] Neumark-Sztainer D, Story M, Hannan PJ, et al. Overweight status and eating patterns among adolescents: where do youth stand in comparison to the Healthy People 2010 Objectives? Am J Public Health 2002;92:844–51.
[22] Neumark-Sztainer D, Wall MM, Story M, et al. Correlates of unhealthy weight-control behaviors among adolescents: implications for prevention programs. Health Psychol 2003;22:88–98.
[23] Pingitore R, Spring B, Garfield D. Gender differences in body satisfaction. Obes Res 1997;5:402–9.
[24] Godin G, Shephard RJ. A simple method to assess exercise behavior in the community. Can J Appl Sport Sci 1985;10:141–6.
[25] Rockett HRH, Breitenbach MA, Frazier AL, et al. Validation of a youth/adolescent food frequency questionnaire. Prev Med 1997;26:808–16.
[26] Rockett HR, Wolf AM, Colditz GA. Development and reproducibility of a food frequency questionnaire to assess diets of older children and adolescents. J Am Diet Assoc 1995;95:336–40.
[27] Lohman T, Roche AF, Martorell R, eds. Anthropometric Standardization Reference Manual. Champaign, IL: Human Kinetics Books, 1988.
[28] Little RJA. Survey nonresponse adjustments for estimates of means. Int Stat Rev 1986;54:137–9.
[29] Crocker P, Sabiston C, Forrestor S, et al. Predicting change in physical activity, dietary restraint, and physique anxiety in adolescent girls. Can J Public Health 2003;94:332–7.
[30] Cooley E, Toray T. Body image and personality predictors of eating disorder symptoms during the college years. Int J Eat Disord 2001;30:28–36.
[31] Stice E, Mazotti L, Krebs M, et al. Predictors of adolescent dieting behaviors: a longitudinal study. Pychol Addict Behav 1998;12:195–205.
[32] Crocker P, Kowalski N, Kowalski K. Smoking behaviour and dietary restraint in young adolescent women: the role of physical self-perceptions. Can J Public Health 2001;92:428–32.
[33] Mikkila V, Lahti-Koski M, Pietinen P, et al. Associates of obesity and weight dissatisfaction among Finnish adolescents. Public Health Nutr 2003;6:49–56.
[34] Fisher M, Schneider M, Pegler C, et al. Eating attitudes, health-risk behaviors, self-esteem, and anxiety among adolescent females in a suburban high school. J Adolesc Health 1991;12:377–84.

[35] Potter BK, Pederson LL, Chan SS, et al. Does a relationship exist between body weight, concerns about weight, and smoking among adolescents? An integration of the literature with an emphasis on gender. Nicotine Tob Res 2004;6:397–425.

[36] Neumark-Sztainer D, Story M, Hannan PJ, et al. Factors associated with changes in physical activity: a cohort study of inactive adolescent girls. Arch Pediatr Adolesc Med 2003;157:803–10.

[37] Sallis JF, Alcaraz J, McKenzie TL, et al. Predictors of change in children's physical activity over 20 months. Am J Prev Med 1999;16:222–9.

[38] McDonald K, Thompson JK. Eating disturbance, body image dissatisfaction, and reasons for exercising: gender differences and correlational findings. Int J Eat Disord 1992;11:289–92.

Journal of Affective Disorders 281 (2021) 767–775

Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad

Research paper

Meaningful Change in Depression Symptoms Assessed with the Patient Health Questionnaire (PHQ-9) and Montgomery-Åsberg Depression Rating Scale (MADRS) Among Patients with Treatment Resistant Depression in Two, Randomized, Double-blind, Active-controlled Trials of Esketamine Nasal Spray Combined With a New Oral Antidepressant

Stacie Hudgens[a,*], Lysbeth Floden[a], Michael Blackowicz[a], Carol Jamieson[b], Vanina Popova[c], Maggie Fedgchin[d], Wayne C. Drevets[e], Kimberly Cooper[f], Rosanne Lane[d], Jaskaran Singh[e,1]

[a] Clinical Outcomes Solutions, Tucson, AZ, US
[b] Janssen Research and Development LLC, Milpitas, CA, US
[c] Janssen Research and Development LLC, Beerse, BE
[d] Janssen Research and Development LLC, Titusville, NJ, US
[e] Janssen Research and Development LLC, San Diego, CA, US
[f] Janssen Research and Development LLC, Spring House, PA, US

ARTICLE INFO

Keywords:
Meaningful change threshold
MADRS
treatment resistant depression
Quality of life
PHQ-9
esketamine

ABSTRACT

Background: Patients with major depressive disorder who do not respond to ≥2 different pharmacological treatments within the current depressive episode are considered to have treatment resistant depression (TRD). This analysis determined meaningful change thresholds (MCT) of the Patient Health Questionnaire (PHQ-9) and Montgomery-Åsberg Depression Rating Scale (MADRS) using anchor-based methods and compared proportions of meaningful changes in patients with TRD across treatment groups from two Phase 3 trials for esketamine nasal spray (SPRAVATO™).

Methods: Data from two Phase 3 trials in patients with TRD, TRANSFORM-1 and -2, were used in this analysis. The MCTs for the PHQ-9 and MADRS were derived using a clinician global impression of severity anchor. Blinded probability density functions displayed score distributions between anchor categories. Proportions of meaningful response were compared between treatment groups using chi-square tests supported by unblinded cumulative distribution functions of change scores.

Results: Baseline scores were similar for the PHQ-9 and MADRS between the esketamine/antidepressant (AD) and AD/placebo groups. The most appropriate MCT on the PHQ-9 was -6 points. By Day 28, 86.5% of patients reached or exceeded the PHQ-9 MCT in the esketamine/AD group compared to 70% in the placebo/AD group. The most appropriate MCT for the MADRS was -10 points. By Day 28, 78.2% of patients reached or exceeded the MADRS MCT in the esketamine/AD group compared to 65.0% in the placebo/AD group.

Conclusions: Individual-level meaningful change for the PHQ-9 and MADRS was effectively quantified using a clinical anchor to interpret efficacy from patients with TRD and their treating clinicians.

Abbreviations: CDF, cumulative distribution function; CGI-S, Clinical Global Impression – Severity; CI, confidence interval; FAP, full analysis population; FAS, full analysis set; MADRS, Montgomery-Åsberg Depression Rating Scale; MCID, minimal clinically important difference; MCT, meaningful change threshold; MDD, major depressive disorder; PDF, probability density function; PHQ-9, Patient Health Questionnaire; PRO, patient reported outcome; SD, standard deviation; SEM, standard error of measurement; SES, standard effect size; TRD, treatment resistant depression.

* Corresponding author at: Strategic Lead Regulatory & Access, Clinical Outcomes Solutions, 1820 E River RD, Suite 220, Tucson, AZ 85718 US.
E-mail address: stacie.hudgens@clinoutsolutions.com (S. Hudgens).
[1] Current affiliation: Neurocrine Biosciences, San Diego, CA, US

https://doi.org/10.1016/j.jad.2020.11.066
Received 21 May 2020; Received in revised form 17 October 2020; Accepted 8 November 2020
Available online 14 November 2020

Introduction

Major depressive disorder (MDD) is associated with excess morbidity and mortality and is a leading cause of disability worldwide in terms of total years lost due to disability (Baldessarini et al., 2017; World Health Organization, Depression Fact Sheet, 2020). About one-third of patients with MDD fail to achieve remission with multiple biogenic amine-based (eg, serotonin, norepinephrine) antidepressant treatments and hence have treatment-resistant depression (TRD) (Fava, 2003). In patients who respond to biogenic amine antidepressants, they remain symptomatic and at risk of self-harm during the typical 3 to 7 weeks to onset of effect (Rush et al., 2006). There is an unmet need to develop treatments providing effective, rapid-acting, and sustained or long-term relief of depressive symptoms, especially in patients with TRD.

Esketamine, the S-enantiomer of ketamine racemate and an N-methyl-d-aspartate receptor antagonist, has been recently approved as a nasal spray for the treatment of TRD (Daly et al., 2018; Fedgchin et al., 2019; Popova et al., 2019; Ochs-Ross et al., 2020).

Patient reported outcome (PRO) measures are effective at capturing the impact of disease on patients' daily lives, including symptom burden, functional abilities, and overall health-related quality of life (FDA, 2009). As these describe patient experience, it is important to provide results framed in meaningful terms. Methods of interpreting patient reported outcomes in terms of the minimal clinically important difference (MCID) were introduced in the late 1980's (Jaeschke et al., 1989). Use of the term MCID was phased out a decade later in favor of the minimal important difference. The purpose of this shift was to remove the focus on what was "clinically important," as the differences in the health-related quality of life are intended to be important to the patient, not the clinician. A second paradigm shift occurred in the mid-2010's when Cappelleri and Bushmakin, 2014 described the important distinction between group-level meaningful difference and individual-level meaningful response. Average differences of change between treatment groups of PRO scores may be statistically significant, but statistical significance also depends on the magnitude of the difference, sample size, and distributional qualities of the change score. Alternatively, patients can be classified as responders based on their change from baseline achieving a meaningful threshold. Around the same time, the Critical Path Institute's PRO Consortium and the Consensus Panel for Outcomes Measurement and Psychometrics: Advancing the Scientific Standards (COMPASS) pushed for preferential use of the term meaningful change threshold (MCT), which describes the threshold of change at which the change becomes meaningful to the patient (Critical Path Institute PRO Consortium, 2016; Hudgens et al., 2016). As a result, the MCT is a more recently used terminology for quantifying patient-level meaningful change.

The preferential method in a clinical trial and regulatory setting for derivation of MCTs is via anchor-based methods estimated as the mean change in a PRO measure in patients who reported an amount of change consistent with a clinically relevant change in global impression of disease severity or health measure with supportive cumulative distribution and probability density function curves (Coon and Cook, 2018; FDA, 2018; FDA, 2019; Wyrwich et al., 2013). In comparison, distribution-based methods use characteristics of the sample, such as one half of the standard deviation, to define an MCT. Although both methods can be used to measure meaningful change, European and United States regulatory agencies prefer anchor-based methods over distribution-based methods and suggest distribution derived thresholds to be used as supportive (Coon and Cook, 2018). In addition to these methods, there are many researcher-made decisions that influence the derivation and choice of an MCT. In most cases, the MCT is derived through a multifaceted triangulation of several approaches that cannot be prescribed in an analysis plan (Revicki et al., 2008).

The primary efficacy evaluation in the two esketamine Phase 3 trials described below was the MADRS, a clinician reported instrument used to measure depression severity. Measurement of patient reported depression symptoms using the 9-item Patient Health Questionnaire (PHQ-9) was also included in these studies to provide additional information to evaluate the efficacy of esketamine in treating TRD.

The objective of this analysis was to assess meaningful change and derive an MCT for the PHQ-9 and MADRS using anchor-based methods and apply the derived MCT to compare meaningful differences in patients in response to the different treatment groups.

Methods

Study design

TRANSFORM-1 and TRANSFORM-2 are two Phase 3, randomized, double-blind, multicenter, active-controlled studies to evaluate the efficacy, safety and tolerability of fixed and flexible doses of esketamine nasal spray (56 mg or 84 mg), respectively, in combination with an oral antidepressant.

In the TRANSFORM-1 study, 346 patients were randomized 1:1:1 to receive double-blind nasal spray treatment with either one of 2 fixed doses of esketamine nasal spray (56 mg or 84 mg) or placebo (Fedgchin et al., 2019). In the TRANSFORM-2 study, 227 patients were randomized 1:1 to receive double-blind nasal spray treatment with flexible dose esketamine (56 mg and 84 mg) or placebo (Popova et al., 2019). In both studies, all patients initiated a new, open-label, oral antidepressant (AD) (duloxetine, escitalopram, sertraline or venlafaxine extended release). The double-blind treatment phase was 4 weeks in duration.

Patients were between 18 and 64 years of age with recurrent MDD (per Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition criteria) or single episode MDD ($\geq$ 2 years), without psychotic features, confirmed by the Mini-International Neuropsychiatric Interview. Patients had moderate to severe depression (Inventory of Depressive Symptomatology total score $\geq$34 and MADRS total score $\geq$28). Patients had received separate therapeutic trials of at least 2 different antidepressants with nonresponse ($\leq$25% improvement) in the current depressive episode at the time of randomization.

The study protocol was reviewed by an institutional review board. Adverse events were reported to the institutional review board per the sponsor's standard operating procedures.

Assessment instruments

9-item patient health questionnaire

This is a 9-item patient reported measure that assesses symptoms of depression and is used to measure the response to treatment (Kroenke et al., 2001). Each item is rated on a 4-point scale, and the responses are summed to provide a total score ranging from 0 to 27 with higher scores indicating a greater frequency of symptoms. The recall period is 2 weeks. Studies have shown the internal consistency and test-retest reliability of the PHQ-9 is excellent (Kroenke et al., 2001; Löwe et al., 2004). Additionally, in a trial setting, the PHQ-9 has been demonstrated to be a valid outcome measure able to accurately assess change in symptom frequency (Al Harbi, 2012). However, to the best of our knowledge, an anchor-based MCT for the PHQ-9 in patients with TRD has not been established. The PHQ-9 was completed at baseline, Day 15, and Day 28.

Montgomery-asberg depression rating scale

This clinician-rated scale is designed to measure depression severity and detects changes due to antidepressant treatments. The scale consists of 10 items, each of which is scored from 0 (symptom not present or normal) to 6 (severe or continuous presence of the symptom), for a total possible score of 60. The MADRS evaluates apparent sadness, reported sadness, inner tension, sleep, appetite, concentration, lassitude, inability to feel (interest level), pessimistic thoughts, and suicidal thoughts. The test exhibits high inter-rater reliability. The typical recall period for the MADRS is 7 days. A higher score indicates a worsening of disease severity. The MADRS was conducted weekly from baseline to Day 28.

Clinical global impression – severity (CGI-S)

This clinician reported measure assesses the global severity of the illness based on the patient's history, psychosocial circumstances, symptoms, behavior, and the impact of the symptoms on the patient's ability to function (Spitzer et al., 1990). This instrument measures the severity of psychopathology in the patient from the clinician's perspective on a scale of zero to 7, with 7 being the most extremely ill patients. A global evaluation is conducted, where a higher score indicates a worsening case of the disease in terms of severity. The CGI-S was captured at various timepoints.

Statistical analysis

PRO analyses were conducted on the full analysis population (FAP) which included patients who received at least one dose of nasal spray study medication and one dose of oral antidepressant in the Double-blind Induction Phase who also had available PRO data on Day 1 (Baseline). For the longitudinal PRO analyses, the entrance criteria were expanded to include patients who had available PRO data at Day 1 (Baseline) and the respective post-baseline target assessment date (eg, Day 15, Day 28). Descriptive statistics were used to characterize patient demographics and clinical characteristics as well as instrument completion rates. All analyses were performed using SAS Version 9.4.

Meaningful change threshold (MCT)

The MCTs for the PHQ-9 and the MADRS were derived based on blinded data from the TRANSFORM-1 study using anchor-based methods and supported with distribution-based methods. The MCTs were then applied to the TRANSFORM-2 study data for the responder analysis.

For the anchor-based analysis, CGI-S was used as an anchor and patients were classified into response groups depending on their level of change in CGI-S over the course of the analysis. Mean (SE) within-patient change from baseline, along with min, max, 95% confidence intervals, and a p-value for the respective paired t-test were calculated for each of the categories of CGI-S change from baseline. In the categories of change, patients were classified among all possible change categories (15 levels, ranging from -7 to 7 where negative changes scores indicate improvement).

The standard effect size (SES) was calculated as the change score mean divided by the standard deviation of the change score (SES of Change). The SES and 95% CI was also calculated for the difference in mean change for each category compared to the adjacent anchor category. Additionally, the SES of Difference in Change was derived as the change score mean divided by the baseline standard deviation, which provided consideration of magnitude change from baseline PHQ-9 scores. Graphic representations of the anchor-based analysis were presented as probability density function (PDF) plots for change in PHQ-9 and MADRS total scores for each anchor group of change in CGI-S at each post-baseline visit using data from TRANSORM-1. These PDFs allow for comparisons of the full distributions of PHQ-9 and MADRS change scores between CGI-S change anchor categories, which provide additional support for selecting the anchor category at which change in PHQ-9 and MADRS is observed.

The PDF plots used the Normal kernel function expressed as:

$$K_0(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \; for - \infty < t < \infty$$

The smoothing approach used the default bandwidth free parameter selection method invoked by the UNIVARIATE procedure in SAS version 9 which minimizes approximate mean integrated standard error, also known as the expected L_2 risk function.

Distribution-based methods, which generate an absolute value for change, rather than directional, were used to supplement the anchor-based derivation of MCT. For the distribution-based MCT estimates, the mean patient data at baseline (Day 1) were used, calculated using ½ of the pooled standard deviation (SD) at baseline and 1 standard error of measurement (SEM), defined as SD x $[1 - \text{reliability}]^{1/2}$, where internal consistency reliability is Cronbach's alpha. Estimates of anchor-based and distribution-based MCTs from TRANSFORM-1 and TRANSFORM-2 were compared on a single graphic to assess the magnitude of meaningful change derived from each method and for comparison across the two trials.

Meaningful change from baseline in score

For both the PHQ-9 and the MADRS, the MCT was determined by evaluating the range of potential values from the categories of change anchor-based results and the distributional based results. Final determination was made through expert, clinical review.

Application of the MCT values to a trial population

The final MCTs for the PHQ-9 and MADRS were applied to data from TRANSFORM-2 in the subsequent unblinded responder analysis. Responder categories were created to classify patients from TRANSFORM-2 as "improved," "declined" or "stable," based on whether a change in the PHQ-9 or MADRS score met or exceeded the MCT. Predefined clinical categories of PHQ-9 and MADRS score were also defined and applied based on a review of the literature. For both the PHQ-9 and MADRS, meaningful change ("improved", "declined" or "stable") was defined as 1-group shift in predefined categories (PHQ-9 categories: Normal (0 to 4); Mild (5 to 9); Moderate (10 to 14); Moderately severe (15 to 19); Severe (≥ 20); MADRS categories: Normal/Symptom Absent (0-6); Mild (7-19); Moderate (20-34); Severe (>34)). Differences in responder status were evaluated using a χ^2 test.

To provide a more robust interpretive context for the responder analysis, cumulative distribution functions (CDFs) for those who improved ("responders") and those who remained stable or declined ("non-responders") were classified using the MADRS scores and plotted by treatment using data from TRANSFORM-2. The CDF analysis complements and extends understanding of responder rates by showing a complete picture of the proportions of patients exceeding thresholds along a continuum of change values between each treatment. The x-axis and y-axis are the change score of the measure and the raw frequency of responders for each change score, respectively.

Results

Baseline demographic and instrument measures (TRANSFORM-1 and TRANSFORM-2)

On average, patients in the FAP for TRANSFORM-2 were mostly middle-aged (mean = 46 years, SD = 11.9), non-Hispanic (93%), white (93%), and female (62%), with a mean of 12 years (SD = 10.2) since diagnosis with TRD (Table 1).

Patient demographics and clinical characteristics were generally similar in TRANSFORM-1 and across treatment groups (Table 1). Completion compliance of the PHQ-9 throughout the course of the analysis was high (> 85%) at each visit from baseline until Day 28 (data not shown) and baseline PHQ-9 total scores were similar in the esketamine nasal spray (56 mg) + AD, esketamine nasal spray (84 mg) + AD, and AD + placebo nasal spray groups (TRANSFORM-1) and for the esketamine nasal spray (56 or 84 mg) + AD and AD + placebo nasal spray treatment groups (TRANSFORM-2).

MCT derivation: PHQ-9

In analyses using combined (blinded) treatment groups, the correlation between change on the CGI-S and change on the PHQ-9 at Day 28 was high (> 0.70) for the categories of change on the CGI-S anchor categories. Categories of change on the CGI-S were used to determine a

Table 1
Summary of Demographic and Clinical Characteristics (FAS; TRANSFORM-1 and TRANSFORM-2)

Characteristic	TRANSFORM-1				TRANSFORM-2		
	Esketamine Nasal Spray (56 mg) + Oral AD N = 115	Esketamine Nasal Spray (84 mg) + Oral AD N = 114	Placebo Nasal Spray + Oral AD N = 113	Total N = 342	Esketamine Nasal Spray (56/84 mg) + Oral AD N = 114	Placebo Nasal Spray + Oral AD N = 109	Total N = 223
Age (years)							
Mean (SD)	46.4 (11.18)	45.7 (11.10)	46.8 (11.36)	46.3 (11.19)	44.9 (12.58)	46.4 (11.14)	45.7 (11.89)
Median	48	47	47	47	45	47	47
Min, max	20, 64	18, 64	18, 64	18, 64	19, 64	20, 64	19, 64
Years since diagnosis							
Mean (SD)	16.1 (11.92)	13.6 (9.90)	15.0 (11.81)	14.9 (11.26)	12.9 (10.35)	11.1 (10.02)	12.0 (10.21)
Median	14	13	12	13	10.5	7	9
Min, max	0, 46	0, 44	0, 49	0, 49	0, 49	0, 50	0, 50
Baseline PHQ-9 Score							
Mean (SD)	20.3 (4.11)	20.7 (3.58)	20.8 (3.69)	20.6 (3.80)	20.2 (3.63)	20.4 (3.73)	20.3 (3.67)
Median	21	21	21	21	20.0	21.0	20.0
Min, max	7, 27	7, 27	12, 27	7, 27	5, 27	10, 27	5, 27
Baseline MADRS Score							
Mean (SD)	37.4 (4.76)	37.8 (5.58)	37.5 (6.16)	37.6 (5.51)	37.0 (5.69)	37.3 (5.66)	37.1 (5.67)
Median	37.0	37.5	37.0	37.0	37.0	37.0	37.0
Min, max	27, 50	25, 51	18, 53	18, 53	22, 48	21, 52	21, 52
Baseline CGI-S*							
Mildly ill	0 (0.0%)	0 (0.0%)	1 (0.3%)	1 (0.3%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Moderately ill	18 (5.3)	24 (7.02)	14 (4.1)	56 (16.4%)	21 (18.6%)	19 (17.4%)	40 (17.9%)
Markedly ill	67 (19.6)	59 (17.3)	70 (20.5)	196 (57.3%)	64 (56.6%)	63 (57.8%)	127 (57.0%)
Severely ill	29 (8.5)	29 (8.5)	25 (7.3)	83 (24.3%)	27 (23.9%)	26 (23.9%)	53 (23.8%)
Among the most extremely ill patients	1 (0.3)	2 (0.6)	3 (0.9)	6 (1.8%)	1 (0.9%)	1 (0.9%)	2 (0.9%)
Sex, n (%)							
Male	34 (29.6%)	35 (30.7%)	32 (28.3%)	101 (29.53%)	39 (34.2%)	46 (42.2%)	85 (38.1%)
Female	81 (70.4%)	79 (69.3%)	81 (71.7%)	241 (70.47%)	75 (65.8%)	63 (57.8%)	138 (61.9%)
Race, n (%)							
White	91 (79.1%)	85 (74.6%)	86 (76.1%)	262 (76.61%)	106 (93.0%)	102 (93.6%)	208 (93.3%)
Black or African American	7 (6.1%)	7 (6.1%)	5 (4.4%)	19 (5.56%)	6 (5.3%)	5 (4.6%)	11 (4.9%)
Asian	2 (1.7%)	1 (0.9%)	2 (1.8%)	5 (1.46%)	1 (0.9%)	1 (0.9%)	2 (0.9%)
American Indian or Alaskan Native	0 (0.0%)	1 (0.9%)	0 (0.0%)	1 (0.29%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Multiple	0 (0.0%)	0 (0.0%)	1 (0.9%)	1 (0.29%)	1 (0.9%)	1 (0.9%)	2 (0.9%)
Other	8 (7.0%)	11 (9.6%)	10 (8.8%)	29 (8.48%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Not Reported	7 (6.1%)	9 (7.9%)	9 (8.0%)	25 (7.31%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Ethnicity, n (%)							
Not Hispanic or Latino	74 (64.3%)	78 (68.4%)	71 (62.8%)	223 (65.20%)	108 (94.7%)	99 (90.8%)	207 (92.8%)
Hispanic or Latino	33 (28.7%)	27 (23.7%)	31 (27.4%)	91 (26.61%)	5 (4.4%)	7 (6.4%)	12 (5.4%)
Not Reported	8 (7.0%)	8 (7.0%)	11 (9.7%)	27 (7.89%)	0 (0.0%)	1 (0.9%)	1 (0.5%)
Unknown	0 (0.0%)	1 (0.9%)	0 (0.0%)	1 (0.29%)	1 (0.9%)	2 (1.8%)	3 (1.3%)

Abbreviations: FAS = full analysis set; Max = maximum; Min = minimum; SD = standard deviation.

* Since depression severity (based on the $IDS\text{-}C_{30}$ and MADRS) was an inclusion criterion for entry in the study, the lowest 3 severity categories of the CGI-S were null and not included in the table.

threshold at which, on average, patients experience minimal meaningful improvement on the PHQ-9 at Day 28 (Table 2). Improvement on the PHQ-9 scale increased monotonically for all the categories from "no change (0)" to "1 category improvement (-1)" to "2 category improvement (-2)" to "3 category improvement (-3). The 95% CIs for mean change on the PHQ-9 within the categories of change anchor categories did not overlap for the 0, -1, and -2 change groups providing evidence of unique groups. Specifically, a one category improvement (-1) is associated with a -6.7 point mean improvement in the PHQ-9 scale score at Day 28 compared to baseline, with a 95% CI [-8.22, -5.24; p<0.0001].

Table 2
Anchor-Based Meaningful Change in PHQ-9 and MADRS from Baseline to Day 28 using the CGI-S[a] (TRANSFORM-1)

Change in CGI-S Anchor Category	CGI-S Correlation	N	Mean (SD)	Min, Max	95% CI of Mean	P-Value[b]	SES of Change[c]	SES of Difference in Change[c]	95% CI of SES Difference
Change in PHQ-9									
Categories of change[d]	0.759								
-6		1	-23.0 (-)	-23, -23	-		-	-	-
-5		9	-21.1 (2.42)	-24, -18	(-22.97, -19.25)	< 0.0001	-8.72	-0.57	(-1.29, 0.16)
-4		45	-18.8 (4.30)	-24, -8	(-20.09, -17.51)	< 0.0001	-4.37	-0.69	(-1.11, -0.28)
-3		51	-15.6 (4.84)	-26, -4	(-16.97, -14.25)	< 0.0001	-3.22	-0.34	(-0.73, 0.04)
-2		58	-14.0 (4.54)	-23, -2	(-15.23, -12.82)	< 0.0001	-3.09	-1.23	(-1.6, -0.86)
-1		81	-6.7 (6.73)	-26, 6	(-8.22, -5.24)	< 0.0001	-1	-0.77	(-1.11, -0.43)
0		61	-2.2 (4.36)	-12, 6	(-3.35, -1.11)	0.0002	-0.51	-	-
1		10	-0.7 (4.22)	-11, 5	(-3.72, 2.32)	0.6124	-0.17	0.35	(-0.33, 1.02)
3		1	0.0 (-)	0, 0	-	-	-	-	-
Change in MADRS									
Categories of change[d]	0.773								
-5		8	-28.3 (7.52)	-39, -18	(-34.53, -21.97)	<.0001	-3.76	0.78	[0.01, 1.54]
-4		45	-34.6 (8.28)	-51, -5	(-37.10, -32.14)	<.0001	-4.19	-0.73	(-1.14, -0.31)
-3		52	-28.5 (8.60)	-49, -10	(-30.69, -26.11)	<.0001	-3.32	-0.82	(-1.2, -0.43)
-2		59	-21.3 (9.07)	-41, 3	(-23.62, -18.89)	<.0001	-2.34	-1.14	(-1.5, -0.78)
-1		80	-10.3 (10.02)	-37, 12	(-12.48, -8.02)	<.0001	-1.02	-0.83	(-1.17, -0.48)
0		61	-2.6 (7.96)	-29, 21	(-4.64, -0.61)	0.0120	-0.33	-	-
1		10	-0.9 (3.07)	-5, 4	(-3.10, 1.30)	0.3783	-0.29	0.23	(-0.44, 0.9)

Abbreviations: CGI-S = Clinician Global Impression of Severity; CI = confidence interval; MADRS = Montgomery-Åsberg Depression Rating Scale; Max = maximum; Min = minimum; PHQ-9 = 9-item Patient Health Questionnaire; SD = standard deviation; SES = standard effect size

[a] The CGI-S evaluates the severity of psychopathology on a scale of 0 to 7, negative change from baseline indicates improvement. Considering total clinical experience, a patient is assessed on severity of mental illness at the time of rating according to: 0 = not assessed, 1 = normal, not at all ill, 2 = borderline mentally ill, 3 = mildly ill, 4 = moderately ill, 5 = markedly ill, 6 = severely ill, and 7 = among the most extremely ill patients.

[b] P-value from paired t-test of mean difference from baseline = 0 with pooled standard deviation.

[c] SES of Change is derived as mean change from baseline divided by the change score standard deviation. SES Difference in Change is derived as the difference in mean change from baseline for each category compared to the adjacent anchor category divided by the baseline standard deviation. Cohen's interpretability parameters for SES are 0.2 (Small), 0.5 (Moderate), and 0.8 (Large).

[d] Categories were only displayed if sample size (N) was ≥ 1.

The difference between the "1 category improvement" and the "no change" groups had a large SES difference in change = -0.77; which is useful for interpreting the magnitude of this within-group change. The mean change in PHQ-9 at each category of change in CGI-S anchor is reported in Table 2. Distribution-based MCT thresholds of the baseline PHQ-9 score were 1.9 (1/2 SD) and 2.1 (SEM).

The blinded PDF plots of change in PHQ-9 stratified by CGI-S anchor change category from TRANSFORM-1 (Fig. 1) demonstrate separation of PDF curves between the CGI-S change categories. The change in PHQ-9 stratified by the categories of change on the CGI-S where substantially greater improvement on the PHQ-9 summary score was observed at 1 point of change on the CGI-S. All the categories of worsening or no change on the CGI-S are centered at 0 change in PHQ-9 score, while all improvement categories of the CGI-S are centered at values of improving on the PHQ-9.

Consideration of the MCT threshold for the PHQ-9 found the -6.7 point mean improvement associated with a 1-point improvement on the CGI-S, together with this body of evidence suggests an appropriate MCT of -6 points for the PHQ-9. The MCT for deterioration could not be derived from this study since the sample size for subjects experiencing deterioration was too small to provide a reliable estimate.

MCT derivation - MADRS

In analyses using combined (blinded) treatment groups, the correlation between change on the CGI-S and change on the MADRS at Day 28 were moderate to high (>0.60). Categories of change on the CGI-S were used to determine a threshold at which, on average, patients experience minimal meaningful improvement on the MADRS at Day 28 (Table 2). Improvement on the MADRS increased monotonically for all the categories from "no change (0)" to "1 category improvement (-1)" to "2 category improvement (-2)" to "3 category improvement (-3) with the 95% CIs for the mean change non-overlapping for all groups with more than 10 patients. Further, a one category improvement (-1) is associated with a 10.3 point mean improvement in the MADRS score at Day 28 compared to baseline (95% CI [-12.48, -8.02]; p<0.0001), and a large SES difference in change = -0.83 (95% CI [-1.17, -0.48]). Distribution-based MCT thresholds of the baseline MADRS score were 2.8 (1/2 SD) and 3.2 (SEM).

The blinded PDF plots of change in MADRS stratified by CGI-S anchor change category from study TRANSFORM-1 (Fig. 1) demonstrate separation of PDF curves between the CGI-S change categories. The change in MADRS stratified by the categories of change on the CGI-S where substantially greater improvement on the MADRS change score was observed at 1 point of improvement on the CGI-S and all other improvement categories. All the categories of worsening or no change on the CGI-S are centered at 0 change in MADRS score.

Consideration of the MCT threshold for the MADRS found a 10.3 point mean improvement associated with a 1-point improvement on the CGI-S, together with this body of evidence suggests an appropriate MCT of -10 points for the MADRS. The MCT for deterioration could not be derived from this study since the sample size for subjects experiencing

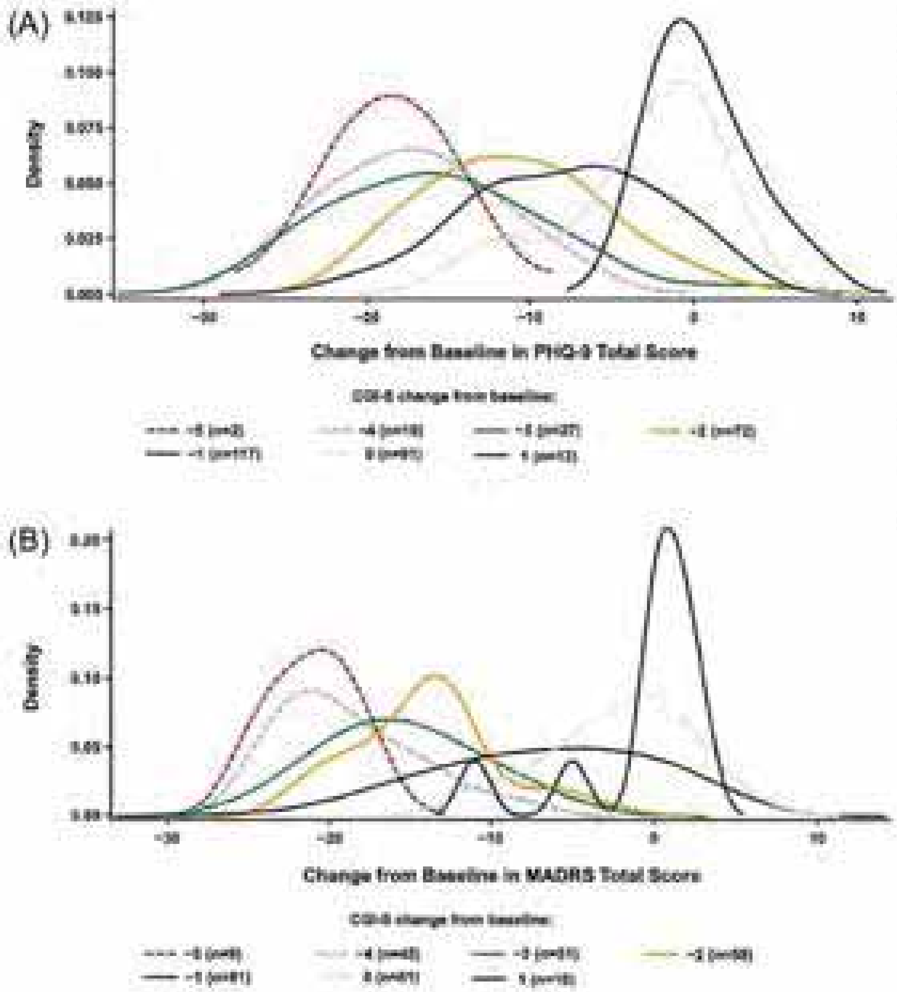


Fig. 1. Blinded Probability Density Function of the Change from Baseline in PHQ-9 and MADRS Total Score at Day 28 Stratified by Categories of Change on the CGI-S from TRANSFORM-1[a]. (A) PHQ-9 Categories of Change on the CGI-S from Baseline. (B) MADRS Categories of Change on the CGI-S from Baseline.

Abbreviations: CGI-S = Clinician Global Impression of Severity; MADRS = Montgomery-Asberg Depression Rating Scale; PHQ-9 = 9-item Patient Health Questionnaire.

[a] The probability density function curves show the distribution of change scores along the x-axis according to the category of anchor change. The y-axis displays the density at a particular point in change. The probably of observing a change score is the integral of the function over the range.

The probability density function plots of change in PHQ-9 and MADRS stratified by CGI-S anchor change category demonstrate separation of the curves between the CGI-S change categories.

deterioration was too small to provide a reliable estimate.

Application of the MCT values to a trial population

Meaningful changes in PHQ-9 from baseline (TRANSFORM-2)

Based on the selected MCT of -6 for the PHQ-9 (TRANSFORM-1), the distribution of response in the TRANSFORM-2 study (improved versus stable versus worsened) from baseline to Day 28 favored the esketamine/AD group over placebo/AD ($p = 0.0071$; Table 3). Specifically, the proportion of patients reaching or exceeding the MCT from baseline to Day 28 was 86.5% in the esketamine/AD group versus 70.0% in the placebo/AD group for a total difference of 16.5 percentage points in favor of esketamine/AD. Results for the predefined clinical categories had higher proportions of responders in each group (90.4% improvers in the esketamine/AD group versus 75.0% of improvers in the placebo/AD group) and were consistent for the between group difference (15.4%) with the results using the MCT. Unblinded CDF plots showed a separation between esketamine/AD group and placebo/AD at the derived MCT of -6 points of change as well as all change scores between -18 and 0 (less than 18-point improvement) for both Days 15 and 28 (Fig. 2).

Meaningful changes in MADRS from baseline (TRANSFORM-2)

Based on the selected MCT of -10 for the MADRS (TRANSFORM-1), the distribution of response in the TRANSFORM-2 study (improved versus stable versus worsened) from baseline to Day 28 favored the esketamine/AD group over placebo/AD ($p = 0.0162$; Table 3). Specifically, the proportion of patients reaching or exceeding the MCT from baseline to Day 28 was 78.2% in the esketamine/AD group versus 65.0% in the placebo/AD group for a total difference of 13.2 percentage points in favor of esketamine/AD. Results for the predefined clinical categories of the MADRS had higher proportions of responders in each group (82.2% improvers in the esketamine/AD group versus 70.0% of improvers in the placebo/AD group) but were consistent for the between group difference (12.2%) with the results using the MCT. Unblinded CDF plots showed a separation between esketamine/AD and placebo/AD distributions at the derived MCT of -10 as well as all change scores between -25 and 0 for Day 15 and between -30 and 0 for Day 28 (Fig. 2).

Discussion

This is the first analysis to identify a meaningful change threshold on the PHQ-9 and the MADRS in patients with TRD using the described anchor-based MCT derivation. An MCT of -6 points was derived as the change in PHQ-9 associated with at least a 1-point improvement on the CGI-S from study TRANSFORM-1. After applying the MCT to data from TRANSFORM-2, the proportions of responders significantly favored esketamine/AD over placebo/AD. This value is supported by the fact that the mean changes are significantly different (Table 2) between "no change" and "1 category improvement" groups with separation of the respective 95% CIs for mean change for all categories. An MCT of -10 was selected and the change in MADRS score using the same methodology and rationale.

Table 3
Categorical Responder Analysis for Change from Baseline in PHQ-9 and MADRS Total Score by Treatment (TRANSFORM-2)

MCT Method	Visit	Response	Oral AD + Nasal Spray Placebo n (%)	Esketamine NasalSpray + Oral AD n (%)	P-value[a]
Change in PHQ-9					
Predefined Categories	Day 15	Improved	70 (67.3%)	84 (75.7%)	0.2913
		Stable	31 (29.8%)	25 (22.5%)	
		Worsened	3 (2.9%)	2 (1.8%)	
	Day 28	Improved	75 (75.0%)	94 (90.4%)	0.0074
		Stable	23 (23.0%)	9 (8.7%)	
		Worsened	2 (2.0%)	1 (1.0%)	
Anchor-based Derivation	Day 15	Improved	57 (54.8%)	74 (66.7%)	0.0932
		Stable	46 (44.2%)	37 (33.3%)	
		Worsened	1 (1.0%)	0 (0.0%)	
	Day 28	Improved	70 (70.0%)	90 (86.5%)	0.0071
		Stable	28 (28.0%)	14 (13.5%)	
		Worsened	2 (2.0%)	0 (0.0%)	
Change in MADRS					
Predefined Categories	Day 15	Improved	55 (53.9%)	74 (69.2%)	0.0748
		Stable	43 (42.2%)	31 (29.0%)	
		Worsened	4 (3.9%)	2 (1.9%)	
	Day 28	Improved	70 (70.0%)	83 (82.2%)	0.0496
		Stable	24 (24.0%)	17 (16.8%)	
		Worsened	6 (6.0%)	1 (1.0%)	
Anchor-based Derivation	Day 15	Improved	43 (42.2%)	58 (54.2%)	0.0484
		Stable	59 (57.8%)	47 (43.9%)	
		Worsened	0 (0.0%)	2 (1.9%)	
	Day 28	Improved	65 (65.0%)	79 (78.2%)	0.0162
		Stable	35 (35.0%)	20 (19.8%)	
		Worsened	0 (0.0%)	2 (2.0%)	

Abbreviations: MADRS = Montgomery-Åsberg Depression Rating Scale; MCT = Meaningful Change Threshold; PHQ-9 = 9-item Patient Health Questionnaire

[a] Fisher's tests were used to test for differences in proportions of categorical change from baseline to Day 15 and Day 28 between treatment and placebo for each domain or summary score.

Primary evidence for MCT derivation of the PHQ-9

The MCT is intended to be the smallest observed change that is considered meaningful to the patient. The smallest categories of change on the CGI-S anchor following "no change" is the step to "minimal improvement," which corresponds to improvement by 1-point on the CGI-S (Fladen et al., 2019). For example, the primary evidence for selecting the PHQ-9 MCT was based on mean change for patients in the 1-point improvement category as well as the range of the upper and lower limit of the 95% confidence intervals (CIs). The observed lower bound of the CI for the 1-point improvement category could not overlap with the upper bound of the CI from the adjacent CGI-S "no change" category. Thus, ensuring the derived MCT was discriminative for patients experiencing actual change on the PHQ-9. It is important to note that this level of categorical change on the CGI-S corresponds with a significant and moderately large magnitude of change on the PHQ-9, as demonstrated by the -0.77 SES difference in change and significant p-value ($p<0.05$).

Methods for MCT derivation

Evidence for an MCT of -6 point on the PHQ-9 was also supported by other, more qualitativeand distribution-based approaches including clinician review. For example, results from the other anchor analyses provide additional insight on selection of the most appropriate MCT. Patients experiencing a 1-point change on the CGI-S score at Day 15 experienced an average change from baseline on the PHQ-9 of -7.9 (95% CI: -8.94, -6.78) points and rounding the lower bound of the 95% CI in the conservative direction would result in an MCT of -6 points (results not shown).

While the distribution-based methods (1/2 SD and SEM) have been considered in legacy MCT methods, these parameters are sensitive to homogeneity in the distribution. More specifically, the patient population in this analysis were enrolled at the severe end of the PHQ-9 distribution with little variability resulting in MCTs less than 3 points to identify a patient as experiencing a meaningful change (approximately 11% change in the total score), which may be too small of a change to be considered meaningful.

Blinded PDFs of the PHQ-9 stratified by categories of change on the CGI-S demonstrated a clear and substantial shift in the distribution of PHQ-9 total score between no change (0) and a 1-point improvement on the CGI-S. The distributions of responses were divided by those who had no change or declined (positive change values) and those who improved. This consistent spacing of the PDF curves supports the use of CGI-S change categories for derivation of the MCT on the PHQ-9.

The MADRS anchor analysis results exhibited similar patterns thus the choice of MCT includes the same justification.

Application of derived MCT (Efficacy)

The proportion of patients with TRD who responded to esketamine/AD according to the patient-relevant MCT between baseline and Day 28 was significantly higher, 16.5 percentage points, compared to patients in the placebo/AD group ($p = 0.0071$). These results are supported by the separation between the unblinded CDFs of PHQ-9 change for esketamine/AD compared to placebo/AD group. The cumulative proportion of responders for most PHQ-9 values of improvement, e.g. between 0 and 16 points, suggests that esketamine/AD was favored over placebo/AD at a wide range of responder thresholds. Similar results on the MADRS CDFs support the difference in responder proportions across a range of MCT values. This is especially important given the understanding that a true definition of meaningful change is more likely a range of values rather than a single threshold.

A limitation of the analysis was that the observed treatment duration was short with only two post-baseline visits at which both the PHQ-9 and MADRS were assessed at the same visit over 28 days of treatment. Furthermore, while the PHQ-9 is a patient reported outcome that captures the patient's perspective, there was no available patient-reported global impression of change or severity to use as an anchor for MCT derivation. In addition, the CGI-S is a global measure based on clinician judgement assessing not only symptoms and behavior, but also the impact of the symptoms on the patient's ability to function. It would be important to confirm the proposed MCTs in future studies where a patient anchor may be available. However, the CGI-S still allowed for an anchor-based assessment using a simple and easily interpreted measure of severity. Another potential limitation of this work was the exclusive focus on differences in responder status and not in remission status.

Despite these limitations, this analysis is the first to offer a measure of meaningful change from the patient's perspective in an especially vulnerable treatment-resistant subpopulation of patients with MDD. The MADRS, used as the primary efficacy evaluation in these two trials, is a well-accepted scale to measure treatment-related differences in improvement of MDD symptoms. However, the MADRS is a clinician-administered measure, whereas the PHQ-9 is completed by the patient which is especially important in MDD where it can be difficult to objectively assess symptom severity and improvement in symptoms from the patient's perspective. The PHQ-9 has largely been used as a clinic-based diagnostic tool for MDD, with less frequent use in clinical trials as a measure of change or improvement in patient-reported symptoms of MDD. Furthermore, the PHQ-9 measures the frequency of the same DSM-5 symptoms for which the MADRS rates severity.

Conclusion

The current analysis is the first to derive an MCT, proposed to be -6, on the PHQ-9 to measure meaningful change from the perspective of the patient using regulatory preferred psychometric anchor-based

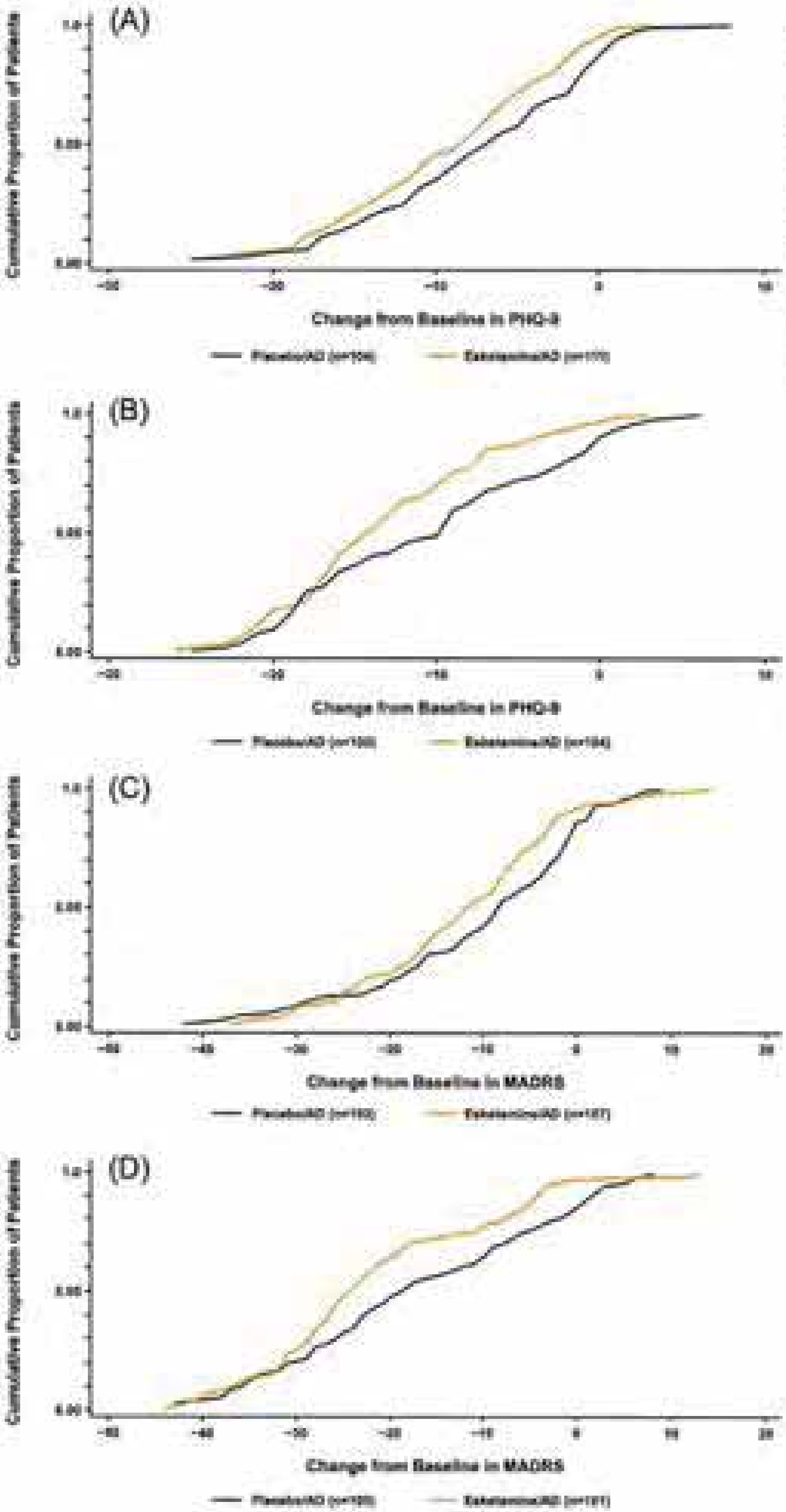


Fig. 2. Unblinded Cumulative Distribution Function Plots for Change from Baseline to Days 15 and 28 on the PHQ-9 and MADRS by Treatment Group (TRANSFORM-2). (A) PHQ-9 Day 15. (B) PHQ-9 Day 28. (C) MADRS Day 15. (D) MADRS Day 28. Abbreviations: CDF = Cumulative Distribution Function; CGI-S = Clinician Global Impression of Severity; MADRS = Montgomery-Åsberg Depression Rating Scale; PHQ-9 = 9-item Patient Health Questionnaire.
CDF plots of meaningful change in PHQ-9 show a separation between esketamine/AD group and placebo/AD at the derived MCT of -6 points of change as well as all change scores between -18 and 0 (less than 18-point improvement) for both Days 15 and 28. CDF plots of the meaningful change in MADRS show a separation between esketamine/AD and placebo/AD distributions at the derived MCT of -10 as well as all change scores between -25 and 0 for Day 15 and between -30 and 0 for Day 28.

methodology. This analysis provides substantial evidence that the PHQ-9, in addition to its utility as a diagnostic tool, can effectively measure change in depression symptoms among patients with TRD. This analysis also proposed the MCT to be -10 on the MADRS. The PHQ-9 should be considered as a key endpoint, alongside the MADRS, in clinical trials of treatment for MDD and TRD to fully characterize symptoms of depression from both a clinician and patient perspective. Furthermore, this analysis supports the positive benefits of esketamine/AD observed in the clinical program from the patient's perspective.

Contributors

All authors contributed to the conception, analysis, interpretation, and revised manuscript content. All authors read and approved the final manuscript.

Funding

This work was supported by Janssen Pharmaceutical Companies of Johnson & Johnson.

Declaration of Competing Interest

CJ, VP, MF, WD, KC, RL, and JS are employees of Janssen Pharmaceutical Companies of Johnson & Johnson. SH, LF, and MB declare no competing interests.

Acknowledgements

Editorial services were provided by Clinical Outcomes Solutions.

References

Al-Harbi, K.S., 2012. Treatment-resistant depression: therapeutic trends, challenges, and future directions. Patient Prefer Adherence 6, 369–388.

Baldessarini, R., Forte, A., Selle, V., Sim, K., Tondo, L., Undurraga, J., Vázquez, G.H., 2017. Morbidity in depressive disorders. Psychother Psychosom 86, 65–72.

Coon, C.D., Cook, K.F., 2018. Moving from significance to real-world meaning: methods for interpreting change in clinical outcome assessment scores. Qual Life Res 27 (1), 33–40.

Critical Path Institute, 2016. PRO Consortium Webinar: Methods for determining clinically meaningful change Available at: https://c-path.org/methods-for-determining-clinically-meaningful-change/. Accessed October 7, 2020.

Daly, E.J., Singh, J.B., Fedgchin, M., Cooper, K., Lim, P., Shelton, R.C., Thase, M.E., Winokur, A., Van Nueten, L., Manji, H., 2018. Efficacy and safety of intranasal esketamine adjunctive to oral antidepressant therapy in treatment-resistant depression: a randomized clinical trial. JAMA Psychiatry. 75(2), 139-148.

Fava, M., 2003. Diagnosis and definition of treatment-resistant depression. Biol Psychiatry 53, 649–659.

Fedgchin, M., Trivedi, M., Daly, E., Melkote, R., Lane, R., Lim, P., Vitagliano, D., Blier, P., 2019. Efficacy and safety of fixed-dose esketamine nasal spray combined with a new oral antidepressant in treatment-resistant depression: results of a randomized, double-blind, active-controlled study (TRANSFORM-1). Int J Neuropsychopharmacol 22 (10), 616–630.

Floden, L., Hodgens, S., Jamieson, C., et al., 2019. Evaluation of depression severity items on the Patient Health Questionnaire (PHQ-9) and Sheehan Disability Scale (SDS) in patients with treatment-resistant depression treated with esketamine nasal spray [poster]. In: Presented at the 32nd Annual Psych Congress. San Diego, CA, US. October 3-6, 2019.

Food and Drug Administration, 2009. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims. Fed Regist 74 (235), 65132–65133.

Food and Drug Administration, 2018. Patient-Focused Drug Development Guidance: Methods to Identify What is Important to Patients and Select, Develop or Modify Fit-for-Purpose Clinical Outcome Assessments. Available at: https://www.fda.gov/drugs/news-events-human-drugs/patient-focused-drug-development-guidance-methods-identify-what-important-patients-and-select. Accessed July 31, 2020.

Food and Drug Administration, 2019. Public Workshop on Patient-Focused Drug Development: Guidance 4 – Incorporating Clinical Outcome Assessments into Endpoints for Regulatory Decision Making. Available at: https://www.fda.gov/drugs/development-approval-process-drugs/public-workshop-patient-focused-drug-development-guidance-4-incorporating-clinical-outcome. Accessed October 8, 2020.

Hodgens, S., Symonds, T., McLeod, L., Coon, C. 2016. Moving the science forward: psychometric considerations and study designs for understanding meaningful change and conducting mixed methods research. In: Presented at the 2016 ISPOR 19th Annual European Congress. Vienna, Austria. October 31, 2016.

Jaeschke, R., Singer, J., Guyatt, G.H., 1989. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials 10 (4), 407–415.

Kroenke, K., Spitzer, R.L., Williams, J.B., 2001. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 16 (9), 606–613.

Löwe, B., Kroenke, K., Herzog, W., Gräfe, K., 2004. Measuring depression outcome with a brief self-report instrument: sensitivity to change of the Patient Health Questionnaire (PHQ-9). J Affect Disord 81 (1), 61–66.

Ochs-Ross, R., Daly, E., Zhang, Y., Lane, R., Lim, P., Morrison, R., Hough, D., Manji, H., Drevets, W., Sanacora, G., Steffens, D., Adler, C., McShane, R., Gaillard, R., Wilkinson, S., Singh, J., 2020. Efficacy and safety of esketamine nasal spray plus an oral antidepressant in elderly patients with treatment-resistant depression: TRANSFORM-3. Am J Geriatr Psychiatry 28 (2), 121–141.

Popova, V., Daly, E., Trivedi, M., Cooper, K., Lane, R., Lim, P., Mazzucco, C., Hough, D., Thase, M., Shelton, R., Molero, P., Vieta, E., Bajbouj, M., Manji, H., Drevets, W., Singh, J., 2019. Efficacy and safety of flexibly dosed esketamine nasal spray combined with a newly initiated oral antidepressant in treatment-resistant depression: a randomized double-blind active-controlled study. Am J Psychiatry 176 (6), 428–438.

Revicki, D., Hays, R.D., Cella, D., Sloan, J., 2008. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol 61 (2), 102–109.

Rush, A., Trivedi, M., Wisniewski, S., Nierenberg, A., Stewart, J., Warden, D., Niederehe, G., Thase, M., Lavori, P., Lebowitz, B., McGrath, P., Rosenbaum, J., Sackeim, H., Kupfer, D., Luther, J., Fava, M., 2006. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. Am J Psychiatry 163, 1905–1917.

Spitzer, R.L., Kroenke, K., Williams, J.B., 1999. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. JAMA 282 (18), 1737–1744.

World Health Organization. Depression Fact Sheet, Updated March 2018. Available at: http://www.who.int/en/news-room/fact-sheets/detail/depression. Accessed April 24, 2020.

Wyrwich, K.W., Norquist, J.M., Lenderking, W.R., Acaster, S., Industry Advisory Committee of International Society for Quality of Life Research (ISOQOL), 2013. Methods for interpreting change over time in patient-reported outcome measures. Qual Life Res 22 (5), 475–483.

nature human behaviour

Registered Report https://doi.org/10.1038/s41562-025-02134-4

Social media use in adolescents with and without mental health conditions

Received: 23 December 2022

Accepted: 11 February 2025

Published online: 5 May 2025

Check for updates

Luisa Fassi[1,2], Amanda M. Ferguson[1], Andrew K. Przybylski[3], Tamsin J. Ford[2] & Amy Orben[1]

Concerns about th… scent
mental health are g…
clinical-level ment… of
how social media u… istered
Report, we analyse… d
11–19 years) includ… gside
quantitative and q…
adolescents with n… time on
social media and w… than
adolescents witho… ences
in social media use…
conditions report… in more
social comparison… mood,
alongside lower h… ower
honest self-disclo… ons only
reported higher time spent. These findings emphasize the need to consider diverse adolescent mental health profiles in policy and clinical practice.

Clinical dx

Adolescents around the world have experienced a decline in their mental health over the past decade[1]. Recent UK data suggests that one in six 7–16-year olds and one in four 17–19-year olds have a probable mental health condition, a clear rise from the one in nine and one in ten recorded in 2017, respectively[2]. As 48% of those with a mental health condition first experience relevant symptoms before the age of 18 years[3], this increased mental health burden will negatively impact society and the economy, as well as adolescent and adult life[4]. Many have raised concerns that this trend has been caused, at least in part, by increased adolescent social media use, which has revolutionized how adolescents live, learn and interact: 95% of 12–17-year olds now report having a social media profile[5].

To address these concerns, academic investigation of social media use and adolescent mental health has increased substantially in recent years[6]. Research teams have recruited adolescent populations in schools, universities or as part of broader community-based samples to identify cross-sectional and longitudinal links between increased smartphone or social media use and scores on questionnaires of depression[7,8], anxiety[9], disordered eating[10] and other mental health symptoms[11–13]. These studies have primarily found small positive associations. Some researchers have used these to argue that there exists a causal link between social media use and mental health declines (that is, "screen time, perhaps especially social media, may have larger effects on adolescent girls' mental health than on boys' and that is indeed what we found, with social media significantly correlated with depressive symptoms […]"[11] p. 13). Such arguments, in turn, have been used to call for restrictive policy regulations to limit smartphone and social media use in adolescent age groups[14].

However, many researchers have also questioned the strength of the current evidence base and highlighted that existing studies do not support the idea that there is a causal relationship linking social media use to mental health. Indeed, the literature provides many conflicting results[15]. Researchers have not only debated about a lack of longitudinal or causal evidence[16], but have also disagreed about what effect sizes matter[17–19] and how to deal with the substantial individual differences present[20,21], which have been linked to factors such as age[22], gender[23,24] and ethnicity[25].

[1]MRC Cognition and Brain Sciences Unit, University of Cambridge, Cambridge, UK. [2]Department of Psychiatry, University of Cambridge, Cambridge, UK. [3]Oxford Internet Institute, University of Oxford, Oxford, UK. e-mail: Luisa.Fassi@mrc-cbu.cam.ac.uk

nature human behaviour

Registered Report

https://doi.org/10.1038/s41562-025-02134-4

Social media use in adolescents with and without mental health conditions

Received: 23 December 2022

Accepted: 11 February 2025

Published online: 5 May 2025

Check for updates

Luisa Fassi[1,2], Amanda M. Ferguson[1], Andrew K. Przybylski[3], Tamsin J. Ford[2] & Amy Orben[1]

Concerns about the relationship between social media use and adolescent mental health are growing, yet few studies focus on adolescents with clinical-level mental health symptoms. This limits our understanding of how social media use varies across mental health profiles. In this Registered Report, we analyse nationally representative UK data (N = 3,340, aged 11–19 years) including diagnostic assessments by clinical raters alongside quantitative and qualitative social media measures. As hypothesized, adolescents with mental health conditions reported spending more time on social media and were less happy about the number of online friends than adolescents without conditions. We also found hypothesized differences in social media use by condition type: adolescents with internalizing conditions reported spending more time on social media, engaging in more social comparison and experiencing greater impact of feedback on mood, alongside lower happiness about the number of online friends and lower honest self-disclosure. In contrast, those with externalizing conditions only reported higher time spent. These findings emphasize the need to consider diverse adolescent mental health profiles in policy and clinical practice.

Adolescents around the world have experienced a decline in their mental health over the past decade[1]. Recent UK data suggests that one in six 7–16-year olds and one in four 17–19-year olds have a probable mental health condition, a clear rise from the one in nine and one in ten recorded in 2017, respectively[2]. As 48% of those with a mental health condition first experience relevant symptoms before the age of 18 years[3], this increased mental health burden will negatively impact society and the economy, as well as adolescent and adult life[4]. Many have raised concerns that this trend has been caused, at least in part, by increased adolescent social media use, which has revolutionized how adolescents live, learn and interact: 93% of 12–17-year olds now report having a social media profile[5].

To address these concerns, academic investigation of social media use and adolescent mental health has increased substantially in recent years[6]. Research teams have recruited adolescent populations in schools, universities or as part of broader community-based samples to identify cross-sectional and longitudinal links between increased smartphone or social media use and scores on questionnaires of depression[10], anxiety[9], disordered eating[11] and other mental health symptoms[12–16]. These studies have primarily found small positive associations. Some researchers have used these to argue that there exists a causal link between social media use and mental health declines (that is, "screen time, perhaps especially social media, may have larger effects on adolescent girls' mental health than on boys' and that is indeed what we found, with social media significantly correlated with depressive symptoms [...]"[17] p. 13). Such arguments, in turn, have been used to call for restrictive policy regulations to limit smartphone and social media use in adolescent age groups[18].

However, many researchers have also questioned the strength of the current evidence base and highlighted that existing studies do not support the idea that there is a causal relationship linking social media use to mental health. Indeed, the literature provides many conflicting results[19]. Researchers have not only debated about a lack of longitudinal or causal evidence[20], but have also disagreed about what effect sizes matter[21–24] and how to deal with the substantial individual differences present[25,26], which have been linked to factors such as age[27], gender[17,28] and ethnicity[29].

[1]MRC Cognition and Brain Sciences Unit, University of Cambridge, Cambridge, UK. [2]Department of Psychiatry, University of Cambridge, Cambridge, UK. [3]Oxford Internet Institute, University of Oxford, Oxford, UK. e-mail: Luisa.Fassi@mrc-cbu.cam.ac.uk

 https://doi.org/10.1038/s41562-025-02134-4

Across these topics of debate, however, researchers have largely overlooked how the kind of instruments used to measure mental health, as well as the populations being studied, limits their ability to draw meaningful inferences about the relationship of social media with adolescent mental health in the first place. So far, most studies have examined school- or community-based adolescent samples[illegible], relating scores on mental health questionnaires (for example, the Hospital Anxiety and Depression Scale[illegible]) to time spent on social media. The rationale for doing so is that questionnaires capturing continuous clinical symptoms are informative when reasoning about social media use in relation to the whole spectrum of mental health, across types of severity and clinical presentation. However, this approach is not a suitable surrogate for studying links between social media and mental health in adolescents with versus without mental health conditions, for two main reasons. First, it reduces the complexity of clinical presentations to the tail end of variation in selected mental health symptoms among mostly healthy individuals. Second, it ignores the potentially important differences between those who endorse symptoms on a questionnaire and those who reach diagnostic criteria in standard clinical classifications. For example, an adolescent can score very highly on a questionnaire measuring depressive symptoms, but not meet the criteria for a diagnosis if the queried symptoms have only been present for a short time or if they are better explained by a different condition or situation.

To address this issue, select studies have moved beyond such an approach, dichotomizing symptom severity by applying cut-off scores to mental health questionnaires to reflect the presence or absence of a mental health diagnosis[27]. However, dichotomization does not solve many of the problems highlighted above and is known to have low sensitivity when predicting clinical diagnoses. Indeed, those with a mental health condition can score below the threshold on some scales[28,29]. In other words, although researchers would like to demonstrate the presence or absence of specific links between mental health conditions and social media use, the measures of psychopathology they employ might not be appropriate for these goals.

Importantly, the assumption that patterns of social media use found in non-clinical or community samples will generalize to those with mental health conditions has not yet been systematically tested. To our knowledge, only a few studies—most qualitative—have documented different social media use experiences in clinical adolescent populations, including those fulfilling stringent diagnostic criteria for a clinical condition, attending mental health services or being hospitalized for suicidal ideation and suicide attempts[illegible]. Adolescents in these studies reported both positive and negative social media experiences, such as enhanced social connection and trouble downregulating their use. Broadly speaking, these experiences aligned with established risk and protective factors previously linked to mental health in offline spaces and suggest there is no clear-cut positive or negative association between mental health and social media use. The studies also raise the idea that vulnerable youth might experience heightened emotional responses to social media use. However, this has not been directly assessed due to the lack of non-clinical comparison groups. Such comparisons are therefore necessary to identify differences in social media use between adolescents with and without mental health conditions.

Owing to the lack of research among young people with mental health conditions[illegible], the important question of whether social media use varies across different types of conditions also remains unaddressed. For example, an adolescent with an internalizing condition (for example, generalized anxiety disorder or depressive disorder) might use and feel impacted by social media differently than an adolescent with an externalizing condition (for example, attention deficit hyperactivity disorder or conduct disorder). This is because, despite both groups presenting mental health symptoms at clinical levels, their experiences of psychopathology can be qualitatively different. Internalizing conditions involve negative emotionality towards the self, expressed through ruminative thought patterns, worries and social withdrawal[illegible]. On the contrary, externalizing conditions involve negative emotionality towards others, expressed through impulsivity, risk taking and disinhibition[illegible]. Studies that assess mental health with select questionnaires cannot comprehensively account for and investigate such clinical diversity. This is a substantial shortcoming, given the need for research to understand how social media use relates to the growing number of adolescents experiencing mental health symptoms at clinical levels.

Table 1 | Summary of our categorization of mental health conditions into internalizing and externalizing

Grouping of mental health conditions	List of mental health conditions diagnosed in the DAWBA
Any	Separation anxiety disorder, generalized anxiety disorder, obsessive-compulsive disorder, specific phobia, social phobia, agoraphobia, panic disorder, post-traumatic stress disorder, other anxiety disorder, major depressive episode, other depressive episode, hyperkinetic disorder, other hyperactivity disorder, oppositional defiant disorder, conduct disorder (general, confined to family, unsocialized, socialized and other), attention deficit hyperactivity disorder, other disruptive behavioural disorders, any behavioural disorder, autism spectrum disorder, eating disorder, tic disorder, selective mutism, psychosis, body dysmorphic disorder, bipolar affective disorder, mania, attachment disorder, feeding disorder, sleep disorder and eliminating disorder
Internalizing	Separation anxiety disorder, generalized anxiety disorder, obsessive-compulsive disorder, specific phobia, social phobia, agoraphobia, panic disorder, post-traumatic stress disorder, other anxiety disorder, body dysmorphic disorder, major depressive episode, other depressive episode and eating disorders
Externalizing	Hyperkinetic disorder, other hyperactivity disorder, oppositional defiant disorder, conduct disorder (general, confined to the family, unsocialized, socialized and other), attention deficit hyperactivity disorder, other disruptive behavioural disorders and any behavioural disorder
Excluded[a] from questions 2 and 3	Autism spectrum disorder, tic disorder, psychotic disorders, mania and bipolar affective disorder

[a]We exclude adolescents diagnosed with these conditions as they do not clearly map onto the symptomatology of either internalizing or externalizing diagnoses[illegible]. Further, because the diagnostic variance of mania and bipolar disorder accounted for by internalizing pathology is lower than most other internalizing disorders, such as anxiety and depression[illegible], we decided to exclude these diagnoses from the internalizing category, together with autism spectrum disorder, tic disorder and psychotic disorders. DAWBA, Development and Wellbeing Assessment.

This Registered Report provides critical data and evidence-based insights into how social media and mental health are related across adolescent populations who meet and do not meet diagnostic criteria for a wide range of mental health conditions. Given the cross-sectional nature of the data and planned analyses, the study results do not provide causal evidence. Hence, all reported coefficients indicate associations, with the possibility of bidirectional relationships and third variables affecting social media use, mental health or the relationship between the two. We analysed the nationally representative Mental Health of Children and Young People (MHCYP) study[illegible], a cross-sectional survey carried out by National Health Service (NHS) Digital in 2017 that collected data from over 3,000 adolescents (11–19 years old) in England. In place of completing self-report measures of mental health, the participants in this study underwent multi-informant standardized diagnostic assessments evaluated by professional clinical raters for different mental health conditions (Table 1). We note that, in the stage 1 report, the terms 'clinical and non-clinical population' were used, but in stage 2 this was changed to 'adolescents with and without mental health conditions' to clarify that participants in this study were not

 https://doi.org/10.1038/s41562-025-02134-4

recruited or diagnosed by a clinic, but instead underwent the mental health assessment as part of the MHCYP study.

Further, to gain a comprehensive understanding of how social media use differs across adolescents with and without a mental health condition, we examined both quantitative and qualitative dimensions of social media use. Measuring only time spent provides a crude and simplistic estimate of social media use, conflating distinct analytical levels and missing a rich range of psychological factors such as appraisal and motivations that might vary as a function of mental health[illegible]. Researchers have therefore called for quantitative time-based measures of social media use to be complemented by more qualitative engagement-based measures capturing adolescents' social media activities and their appraisal of them[illegible]. Such practice is, however, still relatively rare. In this Registered Report, we included both types of measures, namely, time spent on social media and dimensions of social media engagement that could incur mental health risks (that is, online social comparison, monitoring and impact of online feedback and lack of control over time spent online) or benefits (that is, online friendship, as well as opportunities for honest self-disclosure and authentic self-presentation). By complementing quantitative and qualitative dimensions of social media use, this work provides a more solid foundation for mechanistic research aimed at informing future targeted interventions, clinical practice and policy actions benefitting adolescent mental health.

We used existing literature on adolescents' mental health in relation to both online and offline contexts (Table 2) to guide our hypotheses and analyses of the data along three lines of enquiry. Specifically, we evaluated whether social media use differs in adolescents with versus without a mental health condition (Question 1), with an internalizing or externalizing condition versus without a condition (Question 2) and with an internalizing versus externalizing condition (Question 3).

First, we expected adolescents with any mental health condition to report engaging with social media differently than those without a condition. For instance, gathering information about peers is particularly important during adolescence when young people develop a sense of personal and social identity[illegible]. However, high levels of upward social comparison (that is, comparisons with those believed to be of higher status than the self) have been associated with poorer mental health[illegible]. Previous work suggests that most social comparisons made on social media sites are upward rather than downward[illegible], possibly because individuals tend to portray themselves in an ideal manner online[illegible]. Social media could further amplify these processes as platforms offer continuous and more concrete opportunities for comparing oneself with others, such as browsing profiles without initiating social interaction[illegible]. Indeed, engagement in self-denigrating online social comparisons was a common theme raised by adolescent psychiatric inpatients during qualitative interviews[illegible]. We therefore expected adolescents with a mental health condition to engage in more online social comparison than those without a condition (hypothesis H1.1b). Similarly (see Table 2 for a detailed overview of the literature in support of each hypothesis), we expected them to spend more time on social media[illegible] (H1.1a), report more lack of control over time spent online[illegible] (H1.1c), while also monitoring[illegible] (H1.1 d) and feeling more impacted by online feedback[illegible] (H1.1e).

In contrast, we expected adolescents with any condition to engage less with social media in ways that might be protective for their mental health. For instance, we hypothesized that adolescents with a mental health condition are less happy with the number of friends they have online than adolescents without a condition (H1.2f). This is because social connections protect against long-term adverse physical and emotional outcomes, particularly during adolescence[illegible], and young people with mental health conditions often report difficulties with peers, having few friends and wanting to be alone[illegible]. Similarly, we expected adolescents with a mental health condition to engage in less honest online self-disclosure[illegible] (H1.2g) and authentic online self-presentation[illegible] (H1.2h) compared with adolescents without a condition (Table 2).

Second, we predicted social media use to vary between adolescents with different symptomatology, focusing specifically on how the social media use of adolescents with an internalizing or externalizing condition (Table 1) differed from those without a condition.

We know from the mental health literature that internalizing conditions involve negative self-views, rumination, worries and social withdrawal[illegible]. Notably, these symptoms could be relevant to how young people present themselves and engage with others on social media[illegible]. For instance, adolescents with internalizing conditions are more likely to notice discrepancies between their ideal and actual selves[illegible] and may compensate for these discrepancies via impression management offline. We expect this process to also occur online, given the multiple affordances social media platforms offer to curate one's image (for example, deleting old posts and editing new posts). Hence, we hypothesized that adolescents with internalizing conditions would be less likely to engage in authentic self-presentation on social media than adolescents without a mental health condition[illegible] (H2.2h). Further (Table 2), we hypothesized that adolescents with internalizing conditions would spend more time on social media[illegible] (H2.1a), engage in more online social comparison[illegible] (H2.1b) and online feedback monitoring[illegible] (H2.1d), feel more impacted by online feedback[illegible] (H2.1e), be less happy about online friends (H2.2f) and engage in less honest self-disclosure[illegible] (H2.2g) than those without a condition. In contrast, we did not expect a difference in whether they perceive a lack of control over the time they spend online (H2.0c)[illegible].

Compared with internalizing conditions, externalizing conditions involve impulsivity, low self-monitoring and risk taking[illegible]. These symptoms could be reflected in how social media is used by these groups. Hence, we hypothesized that adolescents with externalizing conditions would be more likely than adolescents without a condition to spend more time on social media[illegible] (H2.3a) and perceive they lack control over the time they spend online[illegible] (H2.3c). Further, we expected them to also be more dissatisfied with their number of online friends[illegible] (H2.4f). In contrast, we did not expect differences in the other dimensions of social media use (H2.0b,d,e,g,h; Table 2), since they primarily relate to symptoms of internalizing rather than externalizing conditions.

Our third question examined whether adolescents with internalizing conditions use social media differently than those with externalizing conditions. Specifically, we expected adolescents with internalizing conditions to report engaging in more online social comparison[illegible] (H3.1b) and online feedback monitoring[illegible] (H3.1d), feeling more impacted by online feedback[illegible] (H3.1e), engaging in less honest self-disclosure[illegible] (H3.2g) and less authentic self-presentation on social media[illegible] (H3.2h) than those with externalizing conditions. In contrast, we expected adolescents with externalizing conditions to report having less control over the time they spend online[illegible] compared to adolescents with internalizing conditions (H3.2c). Further, since adolescents with internalizing conditions tend to be unhappy with their social status[illegible] and adolescents with externalizing conditions tend to have trouble making and keeping friends in both online and offline contexts[illegible], we expect both groups to be similarly dissatisfied with their number of online friends (H3.0f). We also hypothesized that both groups would not differ in the amount of time they spend on social media (H3.0a), given that both adolescents with internalizing and externalizing symptoms have been reported to spend more time on social media than adolescents without these symptoms[illegible].

Altogether, this study comprehensively maps and compares different dimensions of social media use in adolescents with and without mental health conditions. Hence, it will lay the foundation for future mechanistic and translational research studying which specific social media dimensions relate to mental health in different adolescent groups. This will be a crucial first step to inform translational research

Table 2 | Review of key social media and mental health literature used to formulate the hypotheses of our study

Social media use	Scientific literature	Relevant hypotheses	Key references
Time spent on social media Item: "When you use social media sites or apps how much time in total do you spend using them on a typical school day/ weekend?"	Research on the relationship between social media use and mental health in different samples has yielded limited and conflicting results. On the one hand, studies have found that young people diagnosed with depression report spending more time on social media compared with non-clinical controls. Further, studies that focus on excessive rather than average time spent on social media show comorbidity between anxiety, depression, attention deficit hyperactivity disorder and excessive time spent on social media. However, in a sample of hospitalized adolescents with psychiatric conditions, the frequency of social media use and perception of overuse was not associated with clinical severity. Additionally, in an independent clinical sample, using social media less, both overall and for messaging, was linked to higher levels of suicidal ideation over the next 30 days. There is more work on community samples available (for example refs. 30,71) examining time spent on social media (mostly self-reported) and its relation to depression, anxiety and other indicators of mental health. Recent reviews and meta-analyses have reached a general agreement that associations are weak and positive (higher social media use is linked with higher levels of anxiety and depression). Overall, despite mixed evidence, research therefore seems to suggest a small relationship between more time spent on social media and lower mental health, both when considering internalizing and externalizing symptoms.	H1.1a: adolescents with any mental health condition will spend more time on social media than adolescents without a condition. H2.1a: adolescents with internalizing mental health conditions will spend more time on social media than adolescents without a condition. H2.2a: adolescents with externalizing mental health conditions will spend more time on social media than adolescents without a condition. H3.0a: adolescents with externalizing mental health conditions will not differ from adolescents with internalizing conditions in time spent on social media.	Valkenburg et al. Cunningham et al. George et al. Gürbüz et al. Fassi et al. Hussain & Griffiths Hamilton et al. Biehm et al. Nesi et al.
Online social comparison Item: "I compare myself to others on social media sites and apps"	In offline contexts, social comparison provides means of gathering information about the social world. This is particularly important during adolescence when young people need to develop a sense of personal and social identity and adjust to bodily changes. Despite social comparison being instrumental for maturation into adulthood, its exacerbation has been associated with poor mental health, particularly in relation to internalizing symptoms. For instance, adolescents with anxiety and eating disorders engage in less favourable and more frequent social comparisons than adolescents without these conditions. Social media platforms offer continuous and more concrete opportunities for social comparison than offline contexts, for instance, by allowing people to browse others' profiles without initiating social interaction. Indeed, adolescents with a depression diagnosis reported unfavourably comparing themselves with others when using social media. Similarly, among adolescents hospitalized for suicidal behaviour, 30% reported engaging in self-denigrating social comparisons, particularly body related ones.	H1.1b: adolescents with any mental health condition will engage in more online social comparison than adolescents without a condition. H2.0b: adolescents with externalizing mental health conditions will not differ in online social comparison from adolescents without a condition. H2.1b: adolescents with internalizing mental health conditions engage in more online social comparison than adolescents without a condition. H3.1b: adolescents with internalizing mental health conditions will engage in more online social comparison than adolescents with externalizing conditions.	Krayer et al. Corning et al. Goodman et al. Rao et al. Thwaites & Dagnan Troop et al. Pempek et al. Radovic et al. Weinstein et al.
Perceived lack of control over time spent online Item: "In general, I spend more time on social media than I mean to"	Difficulties in managing personal goal pursuit in the face of internal, interpersonal and environmental forces that could impede it have been closely linked to the symptomatology of different mental health conditions, particularly externalizing ones. For example, attention deficit hyperactivity disorder and conduct problems are characterized by lower levels of self-monitoring, self-instruction and goal setting. Technological developments now allow individuals to access social media at any location or time of the day. Consequently, online platforms are not limited to a particular environment and repetitions of certain behaviours, such as opening an app or scrolling through one's feed, might result in adolescents feeling unable to reduce the time they spend online despite being motivated to do so. We expect difficulties in self-regulation and goal pursuit to also reflect in how adolescents with mental health conditions engage with social media. Indeed, in a sample of adolescent psychiatric inpatients, self-regulation of social media behaviour was associated with mental health symptoms. Further, in qualitative interviews, 40% of inpatients hospitalized for suicidal behaviour reported trouble regulating social media use and feeling 'addicted'.	H1.1c: adolescents with any mental health condition will be more likely to lack control over time spent online than adolescents without a condition. H2.0c: adolescents with internalizing mental health conditions will not differ in lack control over time spent online from adolescents without a condition. H2.2c: adolescents with externalizing mental health conditions will be more likely to lack control over time spent online than adolescents without a condition. H3.2c: adolescents with externalizing mental health conditions will be more likely to lack control over time spent online than adolescents with internalizing conditions.	Strauman et al. Winds et al. Weinstein et al. Beyer et al.
Monitoring of online feedback Item: "I monitor the amount of likes, comments and shares I get on social media"	Adolescents and adults with depression, anxiety and eating disorders, among other internalizing conditions, have a higher tendency to seek feedback from others, particularly when negative, and engage in more reassurance seeking. Further, compulsivity in social media checking behaviours has been associated with anxiety, depression and problematic smartphone use, therefore becoming a cause of clinical and developmental concern. Consistent with these results, online feedback seeking has been associated with depressive symptoms in adolescent community samples. This association holds even after accounting for the effects of the overall frequency of technology use, offline excessive reassurance seeking and prior depressive symptoms.	H1.1d: adolescents with any mental health condition will be more likely to monitor online feedback than adolescents without a condition. H2.0d: adolescents with externalizing mental health conditions will not differ in monitoring online feedback from adolescents without a condition. H2.1d: adolescents with internalizing mental health conditions will be more likely to monitor online feedback than adolescents without a condition. H3.1d: adolescents with internalizing mental health conditions will be more likely to monitor online feedback than adolescents with externalizing conditions.	Hames et al. Gillet et al. Elhai et al. Barry et al. Nesi & Prinstein Clerkin et al.

https://doi.org/10.1038/s41562-025-02134-4

Nature Human Behaviour | Volume 9 | June 2025 | 1283–1299

Table 2 (continued) | Review of key social media and mental health literature used to formulate the hypotheses of our study

Social media use	Scientific literature	Relevant hypotheses	Key references
Perceived impact of online feedback Item: "The amount of likes, comments and shares I get on social media has an impact on my mood"	Compared with people without a mental health condition, those with internalizing conditions (for example, depression[illegible]) tend to have biased perceptions and recall of interpersonal feedback. This bias is apparent in both their symptoms, such as negative beliefs about the self and worry, as well as neural markers, such as feedback- and error-related negativity[illegible]. Social media platforms offer continuous opportunities for exposure to feedback from peers and strangers. Further, feedback is often more quantifiable and permanent than when received offline (for example, the number of likes received and comments to old posts can be revisited). We therefore expect the biased perception of social feedback demonstrated by some clinical groups to occur and possibly be heightened in online contexts[illegible].	H1.1e: adolescents with any mental health condition will be more likely to feel impacted by online feedback than adolescents without a condition. H2.0e: adolescents with externalizing mental health conditions will not differ in feeling impacted by online feedback from adolescents without a condition. H2.1e: adolescents with internalizing mental health conditions will be more likely to feel impacted by online feedback than adolescents without a condition. H3.1e: adolescents with internalizing mental health conditions will be more likely to feel impacted by online feedback than adolescents with externalizing conditions.	Dobbs[illegible] Tobias & Ito[illegible] Tucker et al.[illegible] Nesi & Prinstein[illegible]
Online friendship Item: "I am happy with the number of friends I have on social media"	In offline contexts, social connections are a protective factor against long-term adverse physical and emotional outcomes, particularly during adolescence[illegible]. In line with this evidence, young people with both internalizing and externalizing mental health conditions report difficulties with peers, having few friends and wanting to be alone[illegible]. Further, adolescents with internalizing conditions tend to be unhappy with their social status and adolescents with externalizing conditions tend to have trouble making and keeping friends, both online and offline[illegible].	H1.2f: adolescents with any mental health condition will be less happy about the number of online friends than adolescents without a condition. H2.2f: adolescents with internalizing mental health condition will be less happy about the number of online friends they have than adolescents without a condition. H2.4f: adolescents with externalizing mental health conditions will be less happy about the number of online friends they have than adolescents without a condition. H3.0f: adolescents with internalizing mental health conditions will be as happy about the number of online friends they have as adolescents with externalizing conditions.	Viner et al.[illegible] Asselmann et al.[illegible] McBride & Preyde[illegible] Sibley et al.[illegible] Bagwell et al.[illegible] Dawson et al.[illegible]
Honest online self-disclosure Item: "I can be honest with people on social media sites and apps about how I am feeling"	Self-disclosure is a communication process by which one person reveals information about themselves to another[illegible]. The extent of self-disclosure has been associated with higher relationship quality, intimacy and well-being in offline contexts. In adolescent psychiatric inpatients, low levels of self-disclosure have been linked to suicidality, with anxiety and depression mediating this association[illegible]. The quality of self-disclosure also differs in people with low compared with high psychological distress. For instance, people that are less distressed tend to disclose more positive information whereas those high in distress tend to disclose more negative and less honest information[illegible]. This effect appears to also occur on social media platforms. For instance, depressed individuals are more likely to post darker, bluer and greyer images than people without depressive symptoms[illegible]. Communicating personal and emotional information increases the risk of embarrassment and rejection[illegible], which is compounded by adolescents' increased sensitivity to peer feedback and anxiety regarding negative social evaluations[illegible]. We expect this to be especially difficult for adolescents with internalizing conditions, as they must balance the rewards associated with self-disclosure[illegible] with considerations of how that disclosure might be received by their peers.	H1.2g: adolescents with any mental health condition will engage in less online honest self-disclosure than adolescents without a condition. H2.0g: adolescents with externalizing mental health conditions will not differ in online honest self-disclosure from adolescents without a condition. H2.2g: adolescents with internalizing mental health conditions will engage in less online honest self-disclosure than adolescents without a condition. H3.2g: adolescents with internalizing mental health conditions will engage in less online honest self-disclosure than adolescents with externalizing conditions.	Sprecher & Hendrick[illegible] Horesh & Apter[illegible] Chen[illegible] Reece & Danforth[illegible] Omarzu[illegible] Van den bos[illegible] Vijayakumar et al.[illegible]
Authentic online self-presentation Item: "My social media profile is a true reflection of myself"	Different internalizing conditions such as anxiety, depression and eating disorders are characterized by the internalization of an ideal self that, once compared with perceptions of the actual self, results in negative self-evaluations. Adolescents with these conditions therefore create perfectionistic self-presentations to combat negative self-narratives and project desirable images of themselves in the mind of others[illegible]. We expect this process to also occur online, given the multiple affordances offered by social media platforms to curate one's image, such as deleting old posts, editing new posts and optimizing messages before sending. Hence, adolescents with a mental health condition, particularly if an internalizing condition, might be more likely to engage in impression management online to compensate for negative self-evaluations[illegible].	H1.2h: adolescents with any mental health condition will engage in less authentic online self-presentation than adolescents without a condition. H2.0h: adolescents with externalizing mental health conditions will not differ in authentic online self-presentation from adolescents without a condition. H2.2h: adolescents with internalizing mental health conditionswill engage in less authentic online self-presentation than adolescents without a condition. H3.2h: adolescents with internalizing mental health conditions will engage in less authentic online self-presentation than adolescents with externalizing conditions.	Flett & Hewitt[illegible] Jain & Sudhir[illegible] O'Connor et al.[illegible] Sassaroli et al.[illegible] Calvo et al.[illegible] Chen[illegible]

The social media items are reported in bold for clarity.

https://doi.org/10.1038/s41562-025-02134-4

https://doi.org/10.1038/s41562-025-02134-4

Table 3 | Descriptive information by mental health condition

Mental health category	N	Age mean	Age s.d.	Sex proportion (male)	SES[a]
No condition	2,821	14.71	2.48	0.50	0.42
Any condition	519	15.10	2.45	0.47	0.41
Externalizing	104	14.27	2.04	0.72	0.45
Internalizing	282	15.64	2.35	0.28	0.40
Other	76	14.00	2.59	0.60	0.36
Between comorbidity	57	13.93	1.80	0.51	0.46

[a]The proportion of participants in the fourth quartile (the most deprived category) based on the Index of Multiple Deprivation. SES, socioeconomic status.

and clinical practice, as well as the design of targeted interventions and policies to improve children's and adolescents' mental health.

Results

Sample description

Our final sample included 3,340 young people (Table 3) aged 11–19 (mean 14.77, s.d. 2.48) years. The sample was 50% male and 50% female. 16% of participants had at least one mental health condition (N = 519). 8% had an internalizing condition (N = 282) and 3% had an externalizing condition (N = 104). Descriptive statistics split by social media items, sex and diagnostic groups are reported in Supplementary Tables 5–8 and Supplementary Figs. 3–5). Further, an overview of our hypotheses and results is provided in Supplementary Table 9.

Any mental health condition versus no condition

We first compared adolescents with versus without mental health conditions, irrespective of condition type (Fig. 1, H1). In line with our hypothesis, we found that adolescents with any mental health condition reported spending more time on social media than adolescents without a condition (H1.1a; g = 0.46 (90% confidence interval (CI) 0.38 to 0.54); NHST (null hypothesis significance testing): β = 0.87, s.e.m. of 0.08, t = 10.48, $P = 2.745 \times 10^{-25}$; EQV (equivalence testing): t(633.8) = 1.21, P = 0.916). Specifically, differences were positive, statistically significant and non-equivalent. That is, we could not reject differences as large or larger than our preregistered equivalence bounds (smallest effect size of interest (SESOI), g = 0.4), indicating potentially meaningful differences between adolescents with and without a condition. On the contrary, the results did not support our hypothesis for online social comparison (H1.1b; g = 0.30 (90% CI 0.22 to 0.39); NHST: β = 0.42, s.e.m. of 0.07, t = 6.52, $P = 7.973 \times 10^{-11}$; EQV: t(641.08) = −1.89, P = 0.026), lack of control over time spent online (H1.1c; g = 0.27 (90% CI 0.19 to 0.35); NHST: β = 0.39, s.e.m. of 0.07, t = 5.52, $P = 3.614 \times 10^{-8}$; EQV: t (674.42) = −2.68, P = 0.002) and impact of online feedback on mood (H1.1e; g = 0.29 (90% CI 0.21 to 0.38); NHST: β = 0.38, s.e.m. of 0.06, t = 6.36, $P = 2.269 \times 10^{-10}$; EQV: t(608.98) = −2.06, P = 0.016). For these dimensions of social media engagement, effect sizes were positive, statistically significant but equivalent (that is, they fell within the preregistered SESOIs). This indicates a difference that is too small to be theoretically meaningful between adolescents with and without a mental health condition. Last, for monitoring of online feedback, we found differences that were not statistically significant and also equivalent, indicating no statistically nor theoretically meaningful differences between adolescents with versus without a condition (H1.1d; g = 0.08 (90% CI −0.01 to 0.15) NHST: β = 0.11, s.e.m. of 0.07, t = 1.55, P = 0.121; EQV: t (653.24) = −6.47, P = 0.000).

We further hypothesized that adolescents with any mental health condition would score lower than adolescents without a condition on dimensions of social media use that could incur mental health benefits, namely, happiness about the number of online friendships (H1.2f), honest self-disclosure (H1.2g) and authentic self-presentation (H1.2h). In line with our hypothesis, we found lower happiness about the number of online friendships (H1.2f; g = −0.37 (90% CI −0.45 to −0.29); NHST: β = −0.33, s.e.m. of 0.04, t = −8.23, $P = 2.660 \times 10^{-16}$; EQV: t(590.4) = 0.56, P = 0.277), for which the effect size was negative, statistically significant and non-equivalent (that is, large enough to be potentially meaningful). In contrast, we did not find differences in honest self-disclosure (H1.2g; g = −0.30 (90% CI −0.38 to −0.22); NHST: β = −0.39, s.e.m. of 0.06, t = −6.37, $P = 2.213 \times 10^{-10}$; EQV: t(629.39) = 1.93, P = 0.028) and authentic self-presentation (H1.2h; g = −0.19 (90% CI −0.28 to −0.11); NHST: β = −0.24, s.e.m. of 0.06, t = −3.98, $P = 7.083 \times 10^{-5}$; EQV: t(624.53) = 4.04, P = 0.000). In these cases, effect sizes were negative and statistically significant but equivalent, suggesting that differences between those with and without a mental health condition were too small to be theoretically meaningful.

Internalizing/externalizing conditions versus no condition

Our second question concerned the extent to which adolescents with internalizing or externalizing conditions differed in their social media use from adolescents without a condition (Fig. 1, H2).

Internalizing versus no condition. Our hypotheses were grounded in the mental health literature, which suggests that anxiety and depressive disorders are characterized by negative self-views, rumination, worries and social withdrawal. We expected these symptoms to be mirrored in adolescents' online experiences. The results supported our hypotheses for time spent on social media (H2.1a; g = 0.62 (90% CI 0.51 to 0.73); NHST: β = 1.12, s.e.m. of 0.11, t = 10.32, $P = 1.609 \times 10^{-23}$; EQV: t(317.39) = 3.21, P = 0.999), online social comparison (H2.1b; g = 0.54 (90% CI 0.43 to 0.65); NHST: β = 0.76, s.e.m. of 0.08, t = 9.12, $P = 1.304 \times 10^{-19}$; EQV: t(318.57) = 2.11, P = 0.994) and the impact of online feedback (H2.1e; g = 0.38 (90% CI 0.27 to 0.49); NHST: β = 0.51, s.e.m. of 0.08, t = 6.61, $P = 4.494 \times 10^{-11}$; EQV: t(306.67) = −0.27, P = 0.385), where we found positive, statistically significant, and non-equivalent effect sizes, suggesting potentially meaningful differences between adolescents with internalizing versus no condition. In contrast, the results did not support our hypothesis for monitoring of online feedback, where differences were not statistically significant and were also too small to be considered meaningful (H2.1d; g = 0.13 (90% CI 0.03 to 0.25); NHST: β = 0.20, s.e.m. of 0.09, t = 2.13, P = 0.033; EQV: t(324.47) = −4.13, P = 0.000).

For those with internalizing conditions, we also hypothesized decreased levels of happiness about the number of online friendships (H2.2f), honest self-disclosure (H2.2g) and authentic self-presentation (H2.2h). Our hypotheses were confirmed for happiness about the number of online friendships (H2.2f; g = −0.45 (90% CI −0.55 to −0.35); NHST: β = −0.40, s.e.m. of 0.05, t = −7.91, $P = 3.49 \times 10^{-15}$; EQV: t(304.45) = −0.69, P = 0.776) and honest self-disclosure (H2.2g; g = −0.31 (90% CI −0.42 to −0.20); NHST: β = −0.41, s.e.m. of 0.08, t = −5.16, $P = 2.670 \times 10^{-7}$; EQV: t(314.2) = 1.32, P = 0.088), where we found negative, statistically significant and potentially meaningful differences. In other words, those with internalizing conditions scored lower than adolescents with no condition. In contrast, we did not find support for meaningful differences in authentic self-presentation (H2.2h; g = −0.19 (90% CI −0.30 to −0.08); NHST: β = −0.25, s.e.m. of 0.08, t = −3.16, P = 0.002; EQV: t(310.64) = 3.08, $P = 0.500 \times 10^{-4}$), where the effect size was statistically significant but equivalent, and therefore too small to be considered meaningful. Last, we expected no differences in lack of control over time spent online for adolescents with internalizing versus no condition (H2.0c). The results did not support our hypothesis, showing positive, statistically significant and potentially meaningful differences (g = 0.43 (90% CI 0.33 to 0.55); NHST: β = 0.60, s.e.m. of 0.09, t = 6.74, $P = 1.91 \times 10^{-11}$; EQV: t(336.39) = 0.499, P = 0.689).

Externalizing versus no condition. Externalizing conditions are characterized by impulsivity, low self-monitoring, and risk taking.

 https://doi.org/10.1038/s41562-025-02134-4

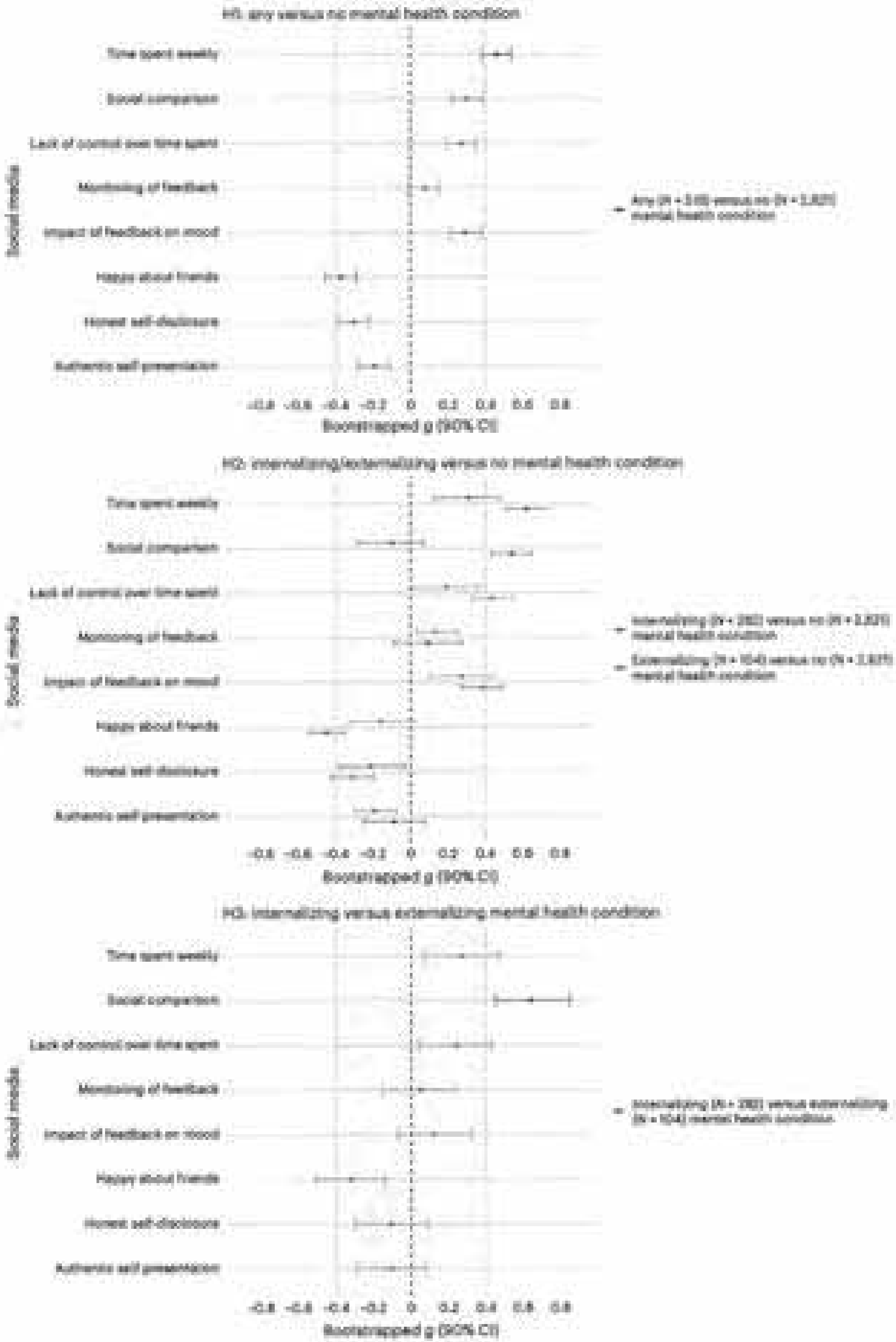


Fig. 1 | Differences in social media use for the three group comparisons. Top: hypothesis 1 (H1, any mental health condition versus no condition). Middle: hypothesis 2 (H2, internalizing/externalizing versus no condition). Bottom: hypothesis 3 (H3, internalizing versus externalizing condition). Data are presented as mean differences based on Hedges's g effect size (g) and its corresponding 90% CI. The shaded area indicates the SESOI (g = 0.4, corresponding to d = 0.4). If the 90% CI lies completely within the SESOI, we concluded equivalence, and therefore no meaningful differences. Bolded effect sizes reflect comparisons that supported our hypotheses, while faded effect sizes reflect comparisons that did not support our hypotheses. The starting sample size includes social media users (N = 519 for adolescents with any mental health condition, N = 282 for internalizing conditions, N = 104 for externalizing conditions and N = 2,821 for no mental health condition). However, the exact sample sizes for each comparison are reported in Supplementary Tables 5 and 6 and differ for each dimension of social media use, given that we planned to analyse these separately.

We expected such symptoms to be reflected in how social media is used by this group. The results supported our hypotheses for time spent on social media (H2.3a: $g = 0.31$ (90% CI 0.13 to 0.48); NHST: $\beta = 0.58$, s.e.m. of 0.17, $t = 3.40$, $P = 0.001$; EQV: $t(105.73) = -0.83$, $P = 0.232$), where we found positive differences that were statistically significant and large enough to be theoretically meaningful. The results did not support our hypotheses for the lack of control over time spent online (H2.3c; $g = 0.19$ (90% CI 0.01 to 0.37); NHST: $\beta = 0.27$, s.e.m. of 0.15, $t = 1.83$, $P = 0.068$; EQV: $t(100.39) = -2.01$, $P = 0.017$) and happiness about the number of online friendships (H2.4f: $g = -0.16$ (90% CI −0.32 to 0.02); NHST: $\beta = -0.12$, s.e.m. of 0.09, $t = -1.39$, $P = 0.165$; EQV: $t(96.65) = 2.34$, $P = 0.023$), where differences were not statistically significant and were too small to be meaningful.

Last, we expected no differences between adolescents with externalizing and no conditions in online social comparison (H2.0b), monitoring of online feedback (H2.0d), feeling impacted by online feedback (H2.0e), honest online self-disclosure (H2.0g) and authentic self-presentation (H2.0h). The results supported our hypotheses for online social comparison (H2.0b: $g = -0.10$ (90% CI −0.28 to 0.07); NHST: $\beta = -0.13$, s.e.m. of 0.14, $t = -0.97$, $P = 0.331$; EQV: $t(104.81) = 2.94$, $P = 0.000$), monitoring of feedback (H2.0d; $g = 0.09$ (90% CI −0.08 to 0.27), NHST: $\beta = 0.13$, s.e.m. of 0.16, $t = 0.82$, $P = 0.410$; EQV: $t(97.07) = -2.92$, $P = 0.004$), honest self-disclosure (H2.0g: $g = -0.21$ (90% CI −0.38 to −0.03), NHST: $\beta = -0.26$, s.e.m. of 0.13, $t = -2.01$, $P = 0.045$; EQV: $t(98.82) = 1.79$, $P = 0.040$), and authentic self-presentation (H2.0h; $g = -0.09$ (90% CI −0.25 to 0.08), NHST: $\beta = -0.11$, s.e.m. of 0.13, $t = -0.85$, $P = 0.395$; EQV: $t(98.29) = 2.96$, $P = 0.003$), as these effect sizes were not statistically significant and were also too small to be considered meaningful. In contrast, the results did not support our hypothesis for the impact of feedback on mood (H2.0e: $g = 0.27$ (90% CI 0.10 to 0.45), NHST: $\beta = 0.34$, s.e.m. of 0.13, $t = 2.71$, $P = 0.007$; EQV: $t(96.78) = -1.14$, $P = 0.120$), where we found positive, significant and potentially meaningful differences.

Internalizing versus externalizing conditions

Our third question focused only on adolescents with a mental health condition and specifically examined whether those with an internalizing condition use social media differently than those with an externalizing condition (Fig. 1, H3).

The results supported our hypotheses for online social comparison (H3.1b; $g = 0.64$ (90% CI 0.45 to 0.85); NHST: $\beta = 0.89$, s.e.m. of 0.16, $t = 5.75$, $P = 0.000$, EQV: $t(203.72) = 2.187$, $P = 0.993$), where we found positive differences that were statistically significant and large enough to be theoretically meaningful in adolescents with internalizing compared with externalizing conditions. In contrast, the results did not support our hypotheses for the monitoring of online feedback (H3.1d; $g = 0.05$ (90% CI −0.15 to 0.24); NHST: $\beta = 0.07$, s.e.m. of 0.18, $t = 0.40$, $P = 0.689$, EQV: $t(157.17) = -2.924$, $P = 0.002$) and impact of online feedback on mood (H3.1e: $g = 0.12$ (90% CI −0.07 to 0.32); NHST: $\beta = 0.16$, s.e.m. of 0.14, $t = 1.12$, $P = 0.261$, EQV: $t(171.21) = -2.421$, $P = 0.008$), where differences were not statistically significant and were too small to be meaningful.

We further hypothesized that adolescents with internalizing conditions would score lower than adolescents with externalizing conditions in lack of control over time spent online (H3.2c), honest self-disclosure (H3.2g) and authentic self-presentation (H3.2h). However, we found neither significant nor meaningful differences across these dimensions (H3.2g: $g = -0.11$ (90% CI −0.30 to 0.09); NHST: $\beta = -0.15$, s.e.m. of 0.15, $t = -0.99$, $P = 0.323$, EQV: $t(170.73) = 2.47$, $P = 0.004$; H3.2h: $g = -0.11$ (90% CI −0.29 to 0.08); NHST: $\beta = -0.14$, s.e.m. of 0.15, $t = -0.93$, $P = 0.351$, EQV: $t(185.78) = 2.51$, $P = 0.005$). Further, the results were inconclusive for lack of control over time spent online (H3.2c; $g = 0.24$ (90% CI 0.04 to 0.43); NHST: $\beta = 0.33$, s.e.m. of 0.17, $t = 1.98$, $P = 0.048$, EQV: $t(156.91) = -1.37$, $P = 0.080$), where we did not find statistically significant differences nor we could reject meaningfully large effect sizes.

Last, we hypothesized that adolescents with internalizing conditions would not differ from adolescents with externalizing conditions in time spent on social media (H3.0a) and happiness in the number of online friendships (H3.0f). The results did not confirm our hypotheses. For time spent on social media, we found positive differences (internalizing higher than externalizing) that were statistically significant and potentially meaningful (H3.0a: $g = 0.27$ (90% CI 0.07 to 0.47); NHST: $\beta = 0.54$, s.e.m. of 0.191, $t = 2.76$, $P = 0.006$, EQV: $t(171.33) = -1.15$, $P = 0.1125$). Further, we found that adolescents with internalizing conditions reported lower happiness about the number of online friends than adolescents with externalizing conditions. In this case, differences were negative, statistically significant and potentially meaningful (H3.0f: $g = -0.32$ (90% CI −0.51 to −0.14); NHST: $\beta = -0.29$, s.e.m. of 0.10, $t = -2.93$, $P = 0.003$, EQV: $t(206.5) = 0.707$, $P = 0.223$).

Exploratory and sensitivity analyses

To extend our findings, we conducted four sets of sensitivity analyses. First, we included adolescents with between-group comorbidities in question 2, such as any internalizing condition with a comorbid externalizing condition, or vice versa, and compared them with those without a condition (Supplementary Tables 15 and 16). Second, we examined the association between mental health severity, conceptualized as the number of conditions (irrespective of diagnostic type), and social media use (Supplementary Table 17). Third, we focused on adolescents with specific conditions and compared their social media use with that of adolescents without a condition. In this case, we only tested conditions for which we were sufficiently powered, namely, major depressive disorder ($N = 86$; Supplementary Table 18), generalized anxiety disorder ($N = 75$; Supplementary Table 19) and social anxiety disorder ($N = 70$; Supplementary Table 20). Last, we tested our hypotheses for time spent on social media separately for school days and weekends, rather than using a composite score (Supplementary Table 21).

Overall, the results were largely in line with our primary findings with a few exceptions. Namely, we found that adolescents with externalizing and between-group comorbidity (Supplementary Table 16) reported less honest self-disclosure than those without a condition. For the sensitivity analysis on mental health severity, we found that the number of conditions, irrespective of type, was associated with time spent on social media, social comparison, monitoring of feedback and impact of feedback on mood (Supplementary Table 17). Last, we did not find differences for time spent on weekdays versus weekends/holidays (Supplementary Table 21). We note that these sensitivity analyses were exploratory and conducted on relatively small sample sizes, which limits the robustness of these findings.

Discussion

In this Registered Report, we analysed differences in social media use between adolescents with and without mental health conditions in a UK sample of over 3,000 participants. Overall, we found significant and meaningful differences across both quantitative (time spent) and qualitative (for example, online social comparison and happiness about the number of online friends) dimensions of social media use.

Interestingly, the largest difference in social media use between those with and without mental health conditions was in the time spent on social media, with the former reporting higher usage. It is important to note that this measure was self-reported, which is known to have only a moderate correlation with objective measures such as sensing data[77]. This raises the question of whether those with mental health conditions perceive that they spend more time on social media or whether they actually do so. Further, we observed that adolescents with a mental health condition reported lower satisfaction with the number of their online friends. In offline contexts, social connections serve as a protective factor against long-term adverse physical and emotional outcomes, especially during adolescence. Our findings therefore suggest that the difficulties with peer relationships experienced by youth

 https://doi.org/10.1038/s41562-025-02134-4

clinical groups offline may also be reflected in their online interactions. For the other dimensions of social media use examined, our results followed the hypothesized direction and were statistically significant (with the exception of monitoring of online feedback). However, the differences were not large enough to be considered meaningful. This might be explained by the relatively high threshold set as our SESOI, which was a moderate effect size grounded in literature on sleep and physical exercise, both established markers of psychopathology.

We next compared social media use between adolescents with a specific mental health condition (internalizing or externalizing) versus no condition. In this case, the results largely supported our hypotheses, whereby adolescents with internalizing conditions demonstrated higher time spent on social media, increased social comparison, greater impact of social media feedback on mood, lower satisfaction with the number of online friends and lower honest self-disclosure compared to those without a mental health condition. Unexpectedly, we also found that adolescents with internalizing conditions reported a higher lack of control over their time spent online, an engagement dimension that we had instead hypothesized would be more pronounced in those with externalizing conditions. For adolescents with externalizing conditions, the only *meaningful* difference when compared with those with no condition was increased time spent online, with no notable differences across other dimensions of social media use. These results might be explained by the fact that the dimensions of social media engagement used in this study were largely framed around internal experiences (that is, they enquired about one's emotions and thoughts) that could be more effective indicators of internalizing rather than externalizing conditions.

Finally, by limiting our focus to only adolescents with mental health conditions, we compared those with internalizing to those with externalizing conditions. We found that adolescents with internalizing conditions engaged in higher online social comparison. This finding aligns with existing research indicating that adolescents with depressive and anxious symptomatology tend to unfavourably compare themselves with others on social media. Given that social media platforms provide continuous and concrete opportunities for social comparison—such as browsing others' profiles without initiating social interaction[illegible]—social comparison may represent a critical mechanism associated with internalizing symptoms on social media. Further, adolescents with internalizing conditions were less happy about the number of their online friends. This may be because their tendency towards negative self-evaluation and social comparison leads them to make negative evaluations of their social status[illegible]. In contrast, adolescents with externalizing conditions might be less focused on social comparison and more on immediate social interactions, resulting in greater satisfaction with their online friendships.

Altogether, this study has three key strengths. First, the complex survey design, which utilized random probability sampling, produced a nationally representative UK sample. Second, all participants in the sample underwent a standardized multi-informant assessment by professional clinical raters. This method provided comprehensive information about participants' mental health without only relying on self-reported questionnaires, self-diagnoses or mental health service use, measures that only capture a subset of individuals with mental health conditions[illegible]. Last, we examined both qualitative and quantitative aspects of social media use, offering insights beyond time spent into relevant engagement dimensions such as online social comparison and the impact of feedback on mood.

We also acknowledge some limitations of our study. First, regarding study design, we analysed raw associations from cross-sectional data. Therefore, no causal or directional inference can be drawn from these findings, including whether the onset of mental health conditions affects the examined dimensions of social media use or vice versa. Further, given the relatively small sample size in our externalizing group (N = 104) and the relatively large number of individuals with between-group comorbidities (N = 57), our findings concerning externalizing conditions should be interpreted with caution. Specifically, our question 2 and 3 results exclude adolescents with a relatively common clinical profile of comorbid externalizing and anxiety disorders. Additional work on how young people with this clinical profile use and experience social media is needed.

Further, although the sample was nationally representative, we cannot determine the extent to which our findings apply outside the UK. Given the socio-cultural factors influencing social media use and mental health conditions, replicating these results in other regions across both the Global North and Global South is crucial if generalizations are to be made[illegible]. We also acknowledge that the data were collected in 2017. While the examined dimensions of social media use remain relevant, the rapid evolution of platforms and user behaviours presents a potential limitation when applying our findings to current trends. Last, we acknowledge limitations regarding our measures. Specifically, we relied on self-reported social media data. Self-reports capture participants' perceptions, such as online social comparison, that cannot be objectively or reliably measured otherwise. However, they often fail to accurately reflect actual patterns in usage, particularly when estimating time spent on social media[illegible].

Future research should aim to replicate and expand on these findings in three key areas. First, studies using experimental and longitudinal designs are essential to clarify the temporal and causal dynamics linking various social media patterns to mental health conditions. Doing so would allow us to disentangle the within-person variation and directional relationship between social media use and mental health symptom presentation, onset or recovery. Second, regarding social media use, future studies could investigate other self-reported and clinically relevant aspects of engagement, such as time displacement from offline activities[illegible], the similarity between perceived self on social media and offline[illegible], as well as goal-directed social media use. Additionally, studies could assess differences in objective social media use, such as time spent on various apps, posting behaviours, active messaging and content exposure. Third, research involving adolescents with intellectual and learning disabilities is necessary to identify differences in this specific clinical group, which was not included in this study.

The results have implications for clinical practice. Specifically, we find key aspects of social media engagement that could inform the creation of guidelines for patient consultations and early intervention strategies. For example, this could include psychoeducation and cognitive-behavioural reappraisal techniques specifically aimed at online social comparison or the impact of social media feedback (for example, 'likes') on mood for adolescents with internalizing conditions.

Over the past years, there has been increasing concern that social media is negatively impacting young people's mental health, but very little research has compared social media use in those with and without mental health conditions[illegible]. In one of the first studies of its kind, we find that young people with mental health conditions report engaging with social media in different ways from those without a condition. This highlights aspects of social media use that might present an increased risk to this already vulnerable group and provides a window for future research to ensure that the digital world is safe for all children regardless of mental health status.

Methods

Ethics information

The MHCYP 2017 survey was reviewed and approved by the West London and GTAC Research Ethics Committee (reference: 16/LO/0155) and the Health Research Authority Confidentiality Advisory Group (reference: 16/CAG/0016) in 2016. Both parents and children provided consent to take part in data collection and were compensated with a £10 voucher for their time. Parents of children under 16 years were interviewed first and permission was sought to interview their child afterwards; the child then provided assent. Conversely, 17–19-year olds were

https://doi.org/10.1038/s41562-025-02134-4

directly asked for their consent, with permission subsequently sought for their parents to be interviewed. Access to the data was granted to the research team by NHS Digital (DARS-NIC-424336-T7K7T-v0.6 r).

Design

The MHCYP study is one of a series of national surveys on the mental health of children and young people in England administered in 1999, 2004, 2017, 2021 and 2022. In this Registered Report, we analysed the 2017 wave collected between January and October 2017: the most recent wave to be made available to researchers as well as the first wave to collect comprehensive data on adolescents' social media use and to include 17–19-year olds. We only analysed data from adolescents who reported being social media users aged 11–19 years, a total of 3,340 participants (50% male and 50% female) out of the full sample of 9,117. The survey was collected using a stratified probability sample of children and young people living in England who were registered with a general practitioner[illegible]. Data were collected via face-to-face interviews with adolescents and their parents. At the same time, if the family agreed, questionnaires were mailed to teachers (for the available data and key demographics see MHCYP 2017[illegible]).

Measures. *Time spent on social media.* The study measured time spent on social media using two questionnaire items: "When you use social media sites or apps, how much time do you spend using them on a typical school day?" (SMTimeSpentS) and "When you use social media sites or apps how much time in total do you spend using them on a typical weekend or holiday day?" (SMTimeSpentW). Participants answered both questions on a nine-point Likert scale: 1, less than 30 min; 2, more than 30 min but less than an hour; 3, 1–2 h; 4, 2–3 h; 5, 3–4 h; 6, 4–5 h; 7, 5–6 h; 8, 6–7 h; 9, more than 7 h. A single variable reflecting average social media hours was created from these two variables (SMTimeSpent; see Supplementary Table 1 for more details). To do so, we first calculated the mean time in hours for each response. For example, if a participant responded "one to two hours", we recoded this as 1.5 h. Participants that responded "more than seven hours" were recoded to 7.5 h, while participants that responded "less than 30 min" were recorded to 15 min (that is, 0.25 h). Weekday hours were then multiplied by 5, and weekend hours were multiplied by 2, and the products were summed and divided by 7 to establish a daily mean social media use variable, measured in hours. The SMTimeSpent variable was coded as continuous[illegible].

In the survey, the questions regarding time spent on social media for both weekdays (SMTimeSpentS) and weekends (SMTimeSpentW) were only asked of adolescents who responded that they use social media sites daily or on most days on a previous questionnaire item (SMFreqofUse: 1, daily or most days). Hence, participants that reported a lower frequency of social media use (that is, SMFreqofUse: 2, a few times a week; 3, once a week; 4, a few times a month; 5, once a month; 6, less often than once a month) were not asked these questions. To handle the resulting missing data in the SMTimeSpent variable, we coded any adolescents who responded to the SMFreqofUse question that they use social media between "a few times a week" and "once a week" to 45 min (0.75 h) and adolescents that responded "a few times a month" and "less often than once a month" to 15 min (0.25 h) on the SMTimeSpent question (for more information about our approach to missing data, see Supplementary Table 2).

Social media engagement. We analysed seven qualitative dimensions of social media engagement measured with questionnaire items tapping into experiences associated with both risks and benefits to adolescent mental health. All measures (summarized with related literature in Table 2) were developed in consultation with a young person advisory group (see the Supplementary Methods for more information), where children and adolescents defined the dimensions of social media use most relevant to them. The measures related to mental health risks encompassed online social comparison ("I compare myself to others on social media"), lack of control over time spent online ("I spend more time on social media than I mean to"), monitoring of online feedback ("I monitor the amount of likes, comments and shares I get on social media") and the impact of online feedback ("The amount of likes, comments and shares I get on social media has an impact on my mood"). On the contrary, the measures indicative of mental health benefits included online friendship ("I am happy with the number of friends I have on social media"), honest self-disclosure ("I can be honest with people on social media sites and apps about how I am feeling) and authentic self-presentation ("My social media profile is a true reflection of myself").

Participants responded to these measures using a five-point Likert scale (1, disagree a lot; 2, disagree a little; 3, neither agree nor disagree; 4, agree a little; 5, agree a lot). We omitted the "Don't know" responses and coded 1–5 responses as continuous, given research suggesting that five-point continuous classifications perform as well as or occasionally better than categorical classifications[illegible]. We performed a sensitivity analysis to test whether examining social media use on weekdays versus weekends/holidays separately, rather than as a weighted average, changed the main results.

Mental health conditions. Face-to-face interviewers completed the Development and Wellbeing Assessment (DAWBA)[illegible] with parents and adolescents aged 11 years or over to establish mental health conditions. During the interview, participants were first led through the 25-item Strengths and Difficulties Questionnaire[illegible] (Supplementary Table 3). Second, the interviewer administered the DAWBA, a diagnostic tool shown to have good validity[illegible] and reliability[illegible]. The DAWBA uses structured and semi-structured questions to assess the presence and severity of symptoms for a wide range of DSM-5 or ICD-10 mental health disorders (Table 1). Each module starts with a few screening items, which, if answered negatively (indicative of the lack of symptoms), allow the interview to proceed to the next module with no loss of accuracy[illegible] (for example, Supplementary Table 4). There is one exception: if a participant scores highly on the Strengths and Difficulties Questionnaire, the interviewer is directed to ask in-depth internalizing disorder DAWBA modules, even if participants screened negative on the initial questions.

After the initial screening items, the subsequent structured items in the DAWBA modules relate directly to diagnostic criteria in DSM-5 and ICD-10. They are close-ended questions about specific mental health symptoms (for example, "In the last 4 weeks, have there been times when you have been very sad, miserable, unhappy or tearful?"). If a participant responds positively to these structured items, they are subsequently asked open-ended questions about these problems (for example, "Please describe your mood—sadness or irritability—and your level of interest in things"). During the assessment, interviewers transcribe open-ended responses verbatim and are also able to add personal comments beneath each response.

A team of clinical raters assessed the DAWBA's structured and qualitative information from all informants to decide whether an adolescent showed evidence of a DSM-5 or ICD-10 mental health condition that would warrant clinical treatment (Supplementary Figs. 1 and 2). To determine the presence of a condition, clinical raters (1) checked that the answers to structured comments were understood by the participants accurately; (2) interpreted any conflicts between child, parent and teacher responses and decided which assessment to prioritize; and (3) identified clinically impairing disorders that did not perfectly fit current operationalized diagnostic criteria or "not otherwise specified diagnosis" such as "other anxiety disorder".

Overall, we coded information on mental health conditions into two separate variables: (1) a binary variable indicating the presence of any mental health condition (diagnosis or no diagnosis) and (2) a categorical variable for the type of condition (internalizing diagnosis,

externalizing diagnosis or no diagnosis). We subdivided conditions identified via the diagnostic assessment into internalizing and externalizing using existing classifications of psychiatric diagnoses[illegible] (Table 1). This distinction draws from transdiagnostic research showing that different conditions (for example, anxiety, depression and eating disorders) are often comorbid, share underlying core symptoms[illegible] and can therefore be grouped to reflect clinical presentations with more validity[illegible]. Further, when we had enough power, we ran additional exploratory analyses examining responses to all social media questions (a–g) separately for individual conditions (specifically, major depressive disorder, generalized anxiety disorder and social anxiety disorder).

Comorbidity data were coded into two separate variables: (1) a binary variable indicating within-group comorbidity (any two diagnoses of either internalizing or externalizing: yes or no, for descriptive purposes only) and (2) a binary variable indicating internalizing–externalizing between-group comorbidity (any comorbid internalizing and externalizing diagnoses: yes or no). Individuals who showed between-group comorbidity were removed before the analysis of questions 2 and 3, given our goal to compare social media use between these groups. To increase the clinical utility of this work, we ran a sensitivity analysis for question 2, including people with between-group comorbidity.

Analysis plan

We conducted all statistical analyses in R version 4.3.1 (R Core Team, 2021), testing the association between time- and engagement-based measures of social media use and mental health conditions using equivalence tests and linear regression models[illegible]. To control for the type I error rate across multiple tests, we set a corrected alpha level of 0.0125, accounting for the false discovery rate across our four tested hypotheses for any given social media item[illegible]. Our analytical approach was based on regressions rather than analysis of variance, as the former allow for more diverse predictors, unbalanced groups and inclusion of covariates in potential exploratory analyses[illegible]. Below, we describe the statistical analyses we used to test our questions and hypotheses (see the Supplementary Information for more details). Further, the analysis code is available on OSF[illegible].

Questions and hypotheses

Question 1: investigate whether adolescents with any mental health condition use social media differently than those without a condition. To address question 1, we tested the association between social media use and mental health conditions. Specifically, we estimated linear regression models with mental health condition as a binary predictor (two levels: diagnosis versus no diagnosis) and social media use as a continuous outcome. The no diagnosis group was set as the reference level for these analyses. Below, we report our null and directional hypotheses, first for dimensions of social media use expected to reflect mental health risks and second for dimensions expected to reflect mental health benefits in adolescents with versus without a condition.

> H0(1.0): adolescents with any mental health condition will not differ from adolescents without a condition in (a) time spent on social media, (b) online social comparison, (c) lacking control over time spent online, (d) monitoring of online feedback, (e) feeling impacted by online feedback, (f) happiness about the number of online friendships, (g) honest online self-disclosure and (h) authentic self-presentation online.

> H1(1.1): adolescents with any mental health condition will score higher than adolescents without a condition in (a) time on social media, (b) online social comparison, (c) lacking control over time spent online, (d) monitoring of online feedback and (e) feeling impacted by online feedback.

> H1(1.2): adolescents with any mental health condition will score lower than adolescents without a condition in (f) happiness about the number of online friendships, (g) honest online self-disclosure and (h) authentic self-presentation online.

To examine whether social media use varies with mental health severity, we conducted sensitivity analyses to test for a linear effect of the number of diagnoses on the social media responses using linear regression models.

Question 2: investigate whether adolescents with an internalizing or externalizing condition use social media differently than those without a condition. After assessing differences in social media use in adolescents with versus without any mental health condition, we examined whether adolescents with internalizing or externalizing conditions use social media differently than adolescents without a condition. Hence, we conducted linear regression models with diagnostic category as a categorical predictor (three levels: internalizing diagnosis, externalizing diagnosis and no diagnosis) and social media use as a continuous outcome. To test our hypotheses, we examined comparisons between two levels of the diagnostic category variable, with no diagnosis set as the reference level. For hypotheses 2.0[i], 2.1 and 2.2, we reported regression coefficients for internalizing versus no diagnosis, while for hypotheses 2.0[e], 2.3 and 2.4 we reported coefficients for externalizing versus no diagnosis. The null hypotheses marked by 'e' or 'i' indicate that, for the considered comparison, the null hypothesis was our primary hypothesis. Hence, for those dimensions of social media use, we expected no difference between adolescents with internalizing or externalizing conditions and those without a condition. We used 'e' to indicate primary null hypotheses related to externalizing versus no condition and 'i' to indicate our primary null hypotheses related to internalising condition.

> H0(2.0): adolescents with internalizing or externalizing conditions will not differ from adolescents without a condition in (a) time on social media, (b) online social comparison[e], (c) lacking control over time spent online[i], (d) monitoring of online feedback[e], (e) feeling impacted by online feedback[e], (f) happiness about the number of online friendships, (g) honest online self-disclosure[e] and (h) authentic self-presentation online[e].

> H1(2.1): adolescents with internalizing condition will score higher than adolescents without a condition in (a) time on social media, (b) online social comparison, (d) monitoring of online feedback and (e) feeling impacted by online feedback.

> H1(2.2): adolescents with internalizing conditions will score lower than adolescents without a condition in (f) happiness about the number of online friendships, (g) honest online self-disclosure and (h) authentic self-presentation online.

> H1(2.3): adolescents with externalizing conditions will score higher than adolescents without a condition in (a) time spent on social media and (c) lack of control over time spent online.

> H1(2.4): adolescents with externalizing conditions will score lower than adolescents without a condition in (f) happiness about the number of online friendships.

Question 3: investigate whether adolescents with an internalizing mental health condition use social media differently than those with an externalizing condition. To address question 3, we examined how adolescents with internalizing conditions differed in social media engagement compared to adolescents with externalizing conditions. To this aim, we compared the internalizing and externalizing

 https://doi.org/10.1038/s41562-025-02134-4

levels of the diagnostic category variable described in question 2, with internalizing as the reference level. Also, in this case, the null hypotheses marked by 'c' indicate our primary hypotheses. Hence, for those dimensions of social media use, we expected no difference between adolescents with internalizing and externalizing conditions.

> H0 (3.0): adolescents with internalizing conditions will not differ from adolescents with externalizing conditions in (a) time on social media[c], (b) online social comparison, (c) lacking control over time spent online, (d) monitoring of online feedback, (e) feeling impacted by online feedback, (f) happiness about the number of online friendships[c], (g) online self-disclosure and (h) authentic self-presentation online.

> H1(3.1): adolescents with internalizing conditions will score higher than adolescents with externalizing conditions in (b) online social comparison, (d) monitoring of online feedback and (e) feeling impacted by online feedback.

> H1(3.2): adolescents with internalizing conditions will score lower than adolescents with externalizing conditions in (c) lack of control over time spent online, (g) online self-disclosure and (h) authentic self-presentation online.

For all regression analyses, we treated the eight dimensions of social media use, including both time- and engagement-based measures, as separate outcomes predicted by information on mental health conditions as the only regressor. While it is common in research to use statistical control to remove confounding effects from a regression coefficient, appropriate control variables should be identified only after justifying a causal structure that includes the outcome, predictors and all theorized confounders. When the selected control variables are inappropriate or remain unjustified, controlling can result in biased regression estimates[illegible]. Further, recent literature warns against controlling for demographic factors such as sex without thought and instead prompts researchers to interrogate how this variable intersects with the predictors and outcomes under investigation[illegible]. In the present work, treating sex or age as a confounding variable would mean ignoring the possibility that there are meaningful sex or age differences in the examined relationships. As our goal is to investigate the overall association between social media use and mental health, we provided a descriptive account of the age and sex of adolescents included in each tested model rather than control for these demographics. For example, in question 1, age and sex of adolescents with any condition versus without a condition are reported for descriptive purposes (Table 3).

Equivalence testing

Given that null hypothesis significance testing does not allow a comprehensive interpretation of statistically non-significant results[illegible], each model was complemented by equivalence tests[illegible] using the bootstrapped two one-sided tests (TOST) with the 'boot_T_TOST' function from the TOSTER package[illegible] to quantify support for the null hypothesis (H0(1–3)). This involves assessing whether the 90% CIs for the effect size lie inside prespecified equivalence bounds that indicate the SESOI. If the CIs lie inside the equivalence bounds, the effect size is interpreted as negligible. On the contrary, if the CIs spread outside the equivalence bounds, the effect size is considered meaningful in size. In this case, the 90% CIs are used rather than 95% because the effect size is tested against two equivalence bounds separately (that is, the upper and lower bound), reflecting $(1 - 2\alpha) \times 100\%$ (see the Supplementary Information for more details).

We established the equivalence bounds by identifying a theoretically meaningful SESOI (Cohen's d = 0.4; refer to the Supplementary Methods for a detailed explanation). After an extensive scoping exercise to identify a suitable theoretical foundation, we determined that the most relevant benchmark for our research questions are everyday behaviours linked to mental health, such as sleep and physical activity. These behaviours, much like social media use, are a regular part of daily routines but, unlike social media use, they are well-established markers of psychopathology, based on both theory[illegible] and matched empirical evidence[illegible]. Consequently, if the actual effect size of social media use is comparable to that observed for behaviours such as sleep and physical activity, we can confidently conclude that social media use also represents an everyday behaviour that exhibits meaningful group-level differences between clinical and non-clinical populations. The interpretation and analysis plan are presented separately for each hypothesis in the Supplementary Information.

Overall, we inferred support for the null hypotheses (that is, no meaningful difference in social media use between groups) if the 90% CI for this association lies within the equivalence bounds. Of note, while a theoretical SESOI based on everyday behaviours and their link to mental health served as our primary effect size of interest to allow for a clear confirmatory approach, we also identified secondary SESOIs based on the effect sizes that are practically and clinically meaningful. These secondary SESOIs play an important role in supporting the interpretation of our findings (as detailed in Supplementary Methods), ensuring the applicability of our study to both academic and practical domains.

Overall, we inferred support for the alternative hypotheses if (1) the coefficient for the association of social media use and our grouping variable for mental health diagnosis was significant, (2) the association followed the hypothesised direction and (3) the CIs for the association did not fall within the equivalence bounds. The interpretation and analysis plan are presented separately for each hypothesis in the Supplementary Information. Further, we detail the planned exploratory analyses in the Supplementary Methods.

Given that not all assumptions of linear models were met (Supplementary Table 9), we complemented our preregistered parametric analyses with exploratory nonparametric tests for all hypotheses (Supplementary Tables 10 and 11). We implemented the Brunner Munzel test based on the 'brunner_munzel' function for NHST and the 'simple_htest' function for equivalence testing from the TOSTER package in R[illegible]. This test, also known as the generalized Wilcoxon test, is a nonparametric test of stochastic equality between two samples that tests against the null hypothesis that for randomly selected values X (that is, social media score in the adolescents without a condition) and Y (that is, social media score in adolescents with a condition), the probability of X being greater than Y is equal to the probability of Y being greater than X. We documented all discrepancies (20% overall) between the parametric and nonparametric test results in Supplementary Tables 12–14.

Sampling plan

Inclusion criteria and sample size. We included all participants in the MHCYP 2017 survey aged between 11 and 19 years at the time of assessment who have answered any of the social media use questions with anything other than "don't know". Only individuals with diagnostic information, including the absence of a diagnosis, were included in the MHCYP dataset. Since we examined each social media use question separately, we included participants who answered at least one of the seven social media engagement questions or the question about time spent on social media. No specific documentation was available regarding the response rates for the social media questions. Given that the questionnaires were administered through interviews with professionals, we expected minimal missing data for the social media responses[illegible]. Hence, using a conservative estimate, we assumed a 5% missingness for our power calculations.

To address question 1, all adolescents with and without a condition were included in our analysis. Hence, in question 1, we included (1) participants with conditions other than internalizing or externalizing disorders (for example, autism spectrum disorder, tic disorder

https://doi.org/10.1038/s41562-025-02134-4

and psychotic disorders; Table 2); (2) participants with within-group comorbidity (multiple internalizing or externalizing diagnoses; for example, comorbid depressive and social anxiety disorder); and (c) participants with between-group comorbidity (both internalizing and externalizing conditions; for example, comorbid depressive and attention deficit hyperactivity disorder).

In contrast, for questions 2 and 3, we included (1) participants with internalizing or externalizing conditions only and (2) participants with within-group comorbidity (multiple internalizing or externalizing diagnoses). Hence, participants with other conditions or between-group comorbidity were excluded from the main analysis, given our goal to compare social media use in adolescents with externalizing and internalizing conditions. However, we ran sensitivity analyses to test the impact of including participants with between-group comorbidities in question 2.

On the basis of the summary demographics[illegible], after accounting for potential missing data, we estimated the sample of complete cases to be around $N = 3{,}854$ (accounting for 5% missingness in the sample of $N = 4{,}057$, 11–19-year olds), approximately 15% of whom received at least one mental health diagnosis ($N = 577$). Available documentation[illegible] suggests that approximately 19.2% of individuals with internalizing conditions have at least one comorbid externalizing condition, and approximately 28% of individuals with externalizing diagnoses have at least one comorbid internalizing diagnosis. This suggests approximately $N = 370$ with internalizing-only diagnoses and $N = 199$ with externalizing-only diagnoses.

Power calculations. We calculated the power by setting the SESOI ($d = 0.4$, refer to Supplementary Methods for a detailed explanation), the alpha level to 0.05 and the estimated sample size (estimated $N = 3{,}854$ based on existing MHCYP documentation). For the equivalence tests, power was calculated using the TOSTER package in R[illegible]. For the regression models, power was determined using the pwr package[illegible] in R. The code for these calculations is available on the OSF[illegible].

H0(1). We calculated the power for equivalence testing to detect a SESOI of $d = 0.4$ given at least $N \geq 577$ individuals with a condition and $N \geq 3{,}277$ individuals without a condition. The results indicate 100% power to reject the presence of effects that are larger than $d = 0.4$.

H1(1). We calculated the power to detect a statistical effect of condition on social media responses using linear regression. The results indicate 100% power to detect the SESOI ($d = 0.4$) with at least $N \geq 577$ individuals with a condition and $N \geq 3{,}277$ individuals without a condition.

H0(2). We calculated the power for equivalence testing to detect a SESOI of $d = 0.4$ given at least $N \geq 370$ individuals with an internalizing only condition and $N \geq 3{,}277$ with no condition. The results indicate 100% power to reject the presence of effects that are larger than $d = 0.4$. We calculated the power for equivalence testing to detect a SESOI of $d = 0.4$ given at least $N > 199$ individuals with an externalizing only condition and $N > 3{,}277$ with no condition. The results indicate 100% power to reject the presence of effects that are larger than $d = 0.4$.

H1(2). We calculated the power to detect a statistical effect of internalizing condition type (internalizing versus no condition) on social media responses using linear regression. The results indicate 100% power to detect the SESOI ($d = 0.4$) with at least $N \geq 370$ individuals with an internalizing-only condition and $N \geq 3{,}277$ with no condition. We calculated power to detect an effect of externalizing diagnosis type (that is, externalizing versus no condition) on social media responses using linear regression. The results indicate 100% power to detect the SESOI ($d = 0.4$) with at least $N \geq 199$ individuals with an externalizing-only condition and $N \geq 3{,}277$ with no condition.

H0(3). We calculated the power for equivalence testing to detect a SESOI of $d = 0.4$ given at least $N \geq 370$ individuals with internalizing only and $N \geq 199$ with externalizing only conditions. The results indicate 96% power to reject the presence of effects that are larger than $d = 0.4$.

H1(3). We calculated the power to detect a statistical effect of internalizing-only versus externalizing-only condition type on social media responses. The results indicate 99.8% power to the SESOI ($d = 0.4$) with at least $N \geq 370$ individuals with internalizing only and $N \geq 199$ with externalizing only conditions.

In addition to our apriori power calculations, we conducted power sensitivity analyses with the final sample size[illegible]. That is, we determined the smallest effect size that is observable with 95% power, given the corrected alpha of 0.0125 and our final sample size (Supplementary Table 22).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The MHCYP dataset is held on behalf of NHS Digital by the UK Data Service. Restrictions apply to the availability of this data for privacy and ethical reasons, which were used under license for this study. Data access can be requested by applying to the Data Access Request Service (DARS; number: DARS-NIC-424336-T7K7T-v0.6). Researchers interested in accessing the data can find further information via the DARS website at https://digital.nhs.uk/services/data-access-request-service-dars/dars-guidance. Source data are provided with this paper.

Code availability

The analysis code can be found on the OSF[illegible].

References

1. World Health Organization. Mental health of adolescents https://www.who.int/news-room/fact-sheets/detail/adolescent-mental-health (2024).
2. NHS. Mental Health of Children and Young People in England in 2022 https://digital.nhs.uk/news/2022/rate-of-mental-disorders-among-17-to-19-year-olds-increased-in-2022-new-report-shows (2022).
3. Paus, T., Keshavan, M. & Giedd, J. N. Why do many psychiatric disorders emerge during adolescence? *Nat. Rev. Neurosci.* **9**, 947–957 (2008).
4. Whiteford, H. A. et al. Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. *Lancet* **382**, 1575–1586 (2013).
5. Ofcom. *Children's Media Use and Attitudes Report* https://www.ofcom.org.uk/__data/assets/pdf_file/0024/234609/childrens-media-use-and-attitudes-report-2022.pdf (2022).
6. Orben, A. Teenagers, screens and social media: a narrative review of reviews and key studies. *Soc. Psychiatry Psychiatr. Epidemiol.* **55**, 407–414 (2020).
7. Yoon, S., Kleinman, M., Mertz, J. & Brannick, M. Is social network site usage related to depression? A meta-analysis of Facebook-depression relations. *J. Affect. Disord.* **248**, 65–72 (2019).
8. Cunningham, S., Hudson, C. C. & Harkness, K. Social media and depression symptoms: a meta-analysis. *Res. Child Adolesc. Psychopathol.* **49**, 241–253 (2021).
9. Woods, H. C. & Scott, H. #Sleepyteens: social media use in adolescence is associated with poor sleep quality, anxiety, depression and low self-esteem. *J. Adolesc.* **51**, 41–49 (2016).
10. Wilksch, S. M., O'Shea, A., Ho, P., Byrne, S. & Wade, T. D. The relationship between social media use and disordered eating in young adolescents. *Int. J. Eat. Disord.* **53**, 96–106 (2020).
11. Beeres, D. T., Andersson, F., Vossen, H. G. M. & Galanti, M. R. Social media and mental health among early adolescents in Sweden: a longitudinal study with 2-year follow-up (KUPOL Study). *J. Adolesc. Health* **68**, 953–960 (2021).

 https://doi.org/10.1038/s41562-025-02134-4

12. Boer, M., Stevens, G., Finkenauer, C., de Looze, M. & van den Eijnden, R. Social media use intensity, social media use problems, and mental health among adolescents: investigating directionality and mediating processes. *Comput. Hum. Behav.* **116**, 106645 (2021).
13. Twenge, J. M., Joiner, T. E., Rogers, M. L. & Martin, G. N. Increases in depressive symptoms, suicide-related outcomes, and suicide rates among US adolescents after 2010 and links to increased new media screen time. *Clin. Psychol. Sci.* **6**, 3–17 (2018).
14. Wait until 8th https://www.waituntil8th.org/ (2024).
15. Valkenburg, P. M., Meier, A. & Beyens, I. Social media use and its impact on adolescent mental health: an umbrella review of the evidence. *Curr. Opin. Psychol.* **44**, 58–68 (2022).
16. Odgers, C. L. & Jensen, M. R. Annual research review: adolescent mental health in the digital age: facts, fears, and future directions. *J. Child Psychol. Psychiatry* **61**, 336–348 (2020).
17. Orben, A. & Przybylski, A. K. The association between adolescent well-being and digital technology use. *Nat. Hum. Behav.* **3**, 173–182 (2019).
18. Twenge, J. M., Haidt, J., Joiner, T. E. & Campbell, W. K. Underestimating digital media harm. *Nat. Hum. Behav.* **4**, 346–348 (2020).
19. Orben, A. & Przybylski, A. K. Reply to: underestimating digital media harm. *Nat. Hum. Behav.* **4**, 349–351 (2020).
20. Beyens, I., Pouwels, J. L., van Driel, I. I., Keijsers, L. & Valkenburg, P. M. The effect of social media on well-being differs from adolescent to adolescent. *Sci. Rep.* **10**, 10763 (2020).
21. Johannes, N., Dienlin, T., Bakhshi, H. & Przybylski, A. K. No effect of different types of media on well-being. *Sci. Rep.* **12**, 61 (2022).
22. Orben, A., Przybylski, A. K., Blakemore, S.-J. & Kievit, R. A. Windows of developmental sensitivity to social media. *Nat. Commun.* **13**, 1649 (2022).
23. Kelly, Y., Zilanawala, A., Booker, C. & Sacker, A. Social media use and adolescent mental health: findings from the UK Millennium Cohort Study. *EClinicalMedicine* **6**, 59–68 (2018).
24. Booker, C. L., Kelly, Y. J. & Sacker, A. Gender differences in the associations between age trends of social media interaction and well-being among 10–15 year olds in the UK. *BMC Public Health* **18**, 321 (2018).
25. Ghai, S., Fassi, L., Awadh, F. & Orben, A. Lack of sample diversity in research on adolescent depression and social media use: a scoping review and meta-analysis. *Clin. Psychol. Sci.* **11**, 759–772 (2023).
26. Fassi, L. et al. Social media use and internalizing symptoms in clinical and community adolescent samples: a systematic review and meta-analysis. *JAMA Pediatr.* **178**, 814–822 (2024).
27. Zigmond, A. S. & Snaith, R. P. The hospital anxiety and depression scale. *Acta Psychiatr. Scand.* **67**, 361–370 (1983).
28. Boyd, J. H. Screening for depression in a community sample: understanding the discrepancies between depression symptom and diagnostic scales. *Arch. Gen. Psychiatry* **39**, 1195 (1982).
29. Manea, L., Gilbody, S. & McMillan, D. Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): a meta-analysis. *Can. Med. Assoc. J.* **184**, E191–E196 (2012).
30. Radovic, A., Gmelin, T., Stein, B. D. & Miller, E. Depressed adolescents' positive and negative use of social media. *J. Adolesc.* **55**, 5–15 (2017).
31. Weinstein, E. et al. Positive and negative uses of social media among adolescents hospitalized for suicidal behavior. *J. Adolesc.* **87**, 63–73 (2021).
32. Nesi, J., Wolff, J. C. & Hunt, J. Patterns of social media use among adolescents who are psychiatrically hospitalized. *J. Am. Acad. Child Adolesc. Psychiatry* **58**, 635–639.e1 (2019).
33. Gurbuz, H., Demir, T., Ozcan, B., Kadak, M. & Poyraz, B. Use of social network sites among depressed adolescents. *Behav. Inf. Technol.* **36**, 517–523 (2017).
34. Eaton, N. R. et al. The structure and predictive validity of the internalizing disorders. *J. Abnorm. Psychol.* **122**, 86–92 (2013).
35. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, 5th edn (American Psychiatric Association, 2022).
36. NHS Digital. *Mental Health of Children and Young People* https://files.digital.nhs.uk/20/1DE800/MHCYP%202017%20Appendix%20F%20-%20Project%20Instructions.pdf (2017).
37. Meier, A. & Reinecke, L. Computer-mediated communication, social media, and mental health: a conceptual and empirical meta-review. *Commun. Res.* **48**, (2020).
38. Verbeij, T., Pouwels, J. L., Beyens, I. & Valkenburg, P. M. The accuracy and validity of self-reported social media use measures among adolescents. *Comput. Hum. Behav. Rep.* **3**, 100090 (2021).
39. Dumas, T. M., Tremblay, P. F., Ellis, W., Millett, G. & Maxwell-Smith, M. A. Does pressure to gain social media attention have consequences for adolescents' friendship closeness and mental health? A longitudinal examination of within-person cross-lagged relations. *Comput. Hum. Behav.* **140**, 107591 (2023).
40. Frison, E. & Eggermont, S. Browsing, posting, and liking on Instagram: the reciprocal relationships between different types of Instagram use and adolescents' depressed mood. *Cyberpsychol. Behav. Soc. Netw.* **20**, 603–609 (2017).
41. Nesi, J., Miller, A. B. & Prinstein, M. J. Adolescents' depressive symptoms and subsequent technology-based interpersonal behaviors: a multi-wave study. *J. Appl. Dev. Psychol.* **51**, 12–19 (2017).
42. Li, J.-B. et al. Associations of intensity and emotional connection related to online social networking use on the risk of incident depression among Chinese adolescents: a prospective cohort study. *J. Affect. Disord.* **308**, 116–122 (2022).
43. Krayer, A., Ingledew, D. K. & Iphofen, R. Social comparison and body image in adolescence: a grounded theory approach. *Health Educ. Res.* **23**, 892–903 (2007).
44. Corning, A. F., Krumm, A. J. & Smitham, L. A. Differential social comparison processes in women with and without eating disorder symptoms. *J. Couns. Psychol.* **53**, 338–349 (2006).
45. Goodman, R. J., Samek, D. R., Wilson, S., Iacono, W. G. & McGue, M. Close relationships and depression: a developmental cascade approach. *Dev. Psychopathol.* **31**, 1451–1465 (2019).
46. Thwaites, R. & Dagnan, D. Moderating variables in the relationship between social comparison and depression: an evolutionary perspective. *Psychol. Psychother. Theory Res. Pract.* **77**, 309–323 (2004).
47. Troop, N. A. Social rank, rank-related life events and eating pathology. *Eur. Eat. Disord. Rev.* **24**, 75–77 (2016).
48. Rao, P. A. et al. Social anxiety disorder in childhood and adolescence: descriptive psychopathology. *Behav. Res. Ther.* **45**, 1181–1191 (2007).
49. Verduyn, P., Gugushvili, N., Massar, K., Täht, K. & Kross, E. Social comparison on social networking sites. *Curr. Opin. Psychol.* **36**, 32–37 (2020).
50. Chou, H.-T. G. & Edge, N. "They are happier and having better lives than I am": the impact of using Facebook on perceptions of others' lives. *Cyberpsychol. Behav. Soc. Netw.* **15**, 117–121 (2012).
51. Gonzales, A. L. & Hancock, J. T. Mirror, mirror on my Facebook wall: effects of exposure to Facebook on self-esteem. *Cyberpsychol. Behav. Soc. Netw.* **14**, 79–83 (2011).
52. Pempek, T. A., Yermolayeva, Y. A. & Calvert, S. L. College students' social networking experiences on Facebook. *J. Appl. Dev. Psychol.* **30**, 227–238 (2009).
53. Strauman, T. J. Self-regulation and psychopathology: toward an integrative translational research paradigm. *Annu. Rev. Clin. Psychol.* **13**, 497–523 (2017).

 https://doi.org/10.1038/s41562-025-02134-4

54. Nesi, J. & Prinstein, M. J. Using social media for social comparison and feedback-seeking: gender and popularity moderate associations with depressive symptoms. *J. Abnorm. Child Psychol.* **43**, 1427–1438 (2015).
55. Gillett, D. A. & Mazza, S. J. Clarifying a construct: an integrative functional model of reassurance-seeking behaviors. *J. Ration. -Emot. Cogn. -Behav. Ther.* **36**, 362–377 (2018).
56. Tobias, M. R. & Ito, T. A. Anxiety increases sensitivity to errors and negative feedback over time. *Biol. Psychol.* **162**, 108092 (2021).
57. Tucker, D. M., Luu, P., Frishkoff, G., Quiring, J. & Poulsen, C. Frontolimbic response to negative feedback in clinical depression. *J. Abnorm. Psychol.* **112**, 667–678 (2003).
58. Asselmann, E., Wittchen, H.-U., Lieb, R. & Beesdo-Baum, K. Sociodemographic, clinical, and functional long-term outcomes in adolescents and young adults with mental disorders. *Acta Psychiatr. Scand.* **137**, 6–17 (2018).
59. McBride, S. & Preyde, M. Loneliness and social isolation in a sample of youth hospitalized for psychiatric illness. *Child Adolesc. Soc. Work J.* **39**, 157–166 (2022).
60. Sibley, M. H., Evans, S. W. & Serpell, Z. N. Social cognition and interpersonal impairment in young adolescents with ADHD. *J. Psychopathol. Behav. Assess.* **32**, 193–202 (2010).
61. Horesh, N. & Apter, A. Self-disclosure, depression, anxiety, and suicidal behavior in adolescent psychiatric inpatients. *Crisis* **27**, 66–71 (2006).
62. Chen, H. Antecedents of positive self-disclosure online: an empirical study of US college students' Facebook usage. *Psychol. Res. Behav. Manag.* **10**, 147–153 (2017).
63. Flett, G. L. & Hewitt, P. L. in *Social Anxiety*, 159–187 (Academic Press, 2014).
64. Jain, M. & Sudhir, P. M. Dimensions of perfectionism and perfectionistic self-presentation in social phobia. *Asian J. Psychiatry* **3**, 216–221 (2010).
65. Krueger, R. F., McGue, M. & Iacono, W. G. The higher-order structure of common DSM mental disorders: internalization, externalization, and their connections to personality. *Personal. Individ. Differ.* **30**, 1245–1259 (2001).
66. Nesi, J. et al. Selfie appearance investment and peer feedback concern: multimethod investigation of adolescent selfie practices and adjustment. *Psychol. Pop. Media* **10**, 488–499 (2021).
67. Higgins, E. T. Self-discrepancy: a theory relating self and affect. *Psychol. Rev.* **94**, 319–340 (1987).
68. Sassaroli, S. et al. Perfectionism in depression, obsessive-compulsive disorder and eating disorders. *Behav. Res. Ther.* **46**, 757–765 (2008).
69. Calvo, R. et al. Obsessive–compulsive personality disorder traits and personality dimensions in parents of children with obsessive–compulsive disorder. *Eur. Psychiatry* **24**, 201–206 (2009).
70. O'Connor, R. C., Rasmussen, S. & Hawton, K. Predicting depression, anxiety and self-harm in adolescents: the role of perfectionism and acute life stress. *Behav. Res. Ther.* **48**, 52–59 (2010).
71. George, M. J. et al. Young adolescents' digital technology use, perceived impairments, and well-being in a representative sample. *J. Pediatr.* **219**, 180–187 (2020).
72. Hussain, Z. & Griffiths, M. D. Problematic social networking site use and comorbid psychiatric disorders: a systematic review of recent large-scale studies. *Front. Psychiatry* **9**, 686 (2018).
73. Dawson, A. E., Wymbs, B. T., Evans, S. W. & DuPaul, G. J. Exploring how adolescents with ADHD use and interact with technology. *J. Adolesc.* **71**, 119–137 (2019).
74. Gilbert, J. A. et al. Converging evidence for enduring perceptions of low social status in individuals in remission from depression. *J. Affect. Disord.* **294**, 661–670 (2021).
75. Bagwell, C. L., Molina, B. S. G., Pelham, W. E. & Hoza, B. Attention-deficit hyperactivity disorder and problems in peer relations: predictions from childhood to adolescence. *J. Am. Acad. Child Adolesc. Psychiatry* **40**, 1285–1292 (2001).
76. Barry, C. T., Sidoti, C. L., Briggs, S. M., Reiter, S. R. & Lindsey, R. A. Adolescent social media use and mental health from adolescent and parent perspectives. *J. Adolesc.* **61**, 1–11 (2017).
77. Parry, D. A. et al. A systematic review and meta-analysis of discrepancies between logged and self-reported digital media use. *Nat. Hum. Behav.* **5**, 1535–1547 (2021).
78. Allen, N. B. & Badcock, P. B. T. The social risk hypothesis of depressed mood: evolutionary, psychosocial, and neurobiological perspectives. *Psychol. Bull.* **129**, 887–913 (2003).
79. Stuart, A. L. et al. Comparison of self-report and structured clinical interview in the identification of depression. *Compr. Psychiatry* **55**, 866–869 (2014).
80. Mathai, J., Anderson, P. & Bourne, A. Comparing psychiatric diagnoses generated by the strengths and difficulties questionnaire with diagnoses made by clinicians. *Aust. N. Z. J. Psychiatry* **38**, 639–643 (2004).
81. Valkenburg, P. M. & Peter, J. Social consequences of the internet for adolescents: a decade of research. *Curr. Dir. Psychol. Sci.* **18**, 1–5 (2009).
82. Bunker, C. J. & Kwan, V. S. Y. Similarity between perceived selves on social media and offline and its relationship with psychological well-being in early and late adulthood. *Comput. Hum. Behav.* **152**, 108025 (2024).
83. NHS. Mental Health of Children and Young People in England, 2017 https://digital.nhs.uk/data-and-information/publications/statistical/mental-health-of-children-and-young-people-in-england/2017/2017 (2017).
84. Rhemtulla, M., Brosseau-Liard, P. É. & Savalei, V. When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychol. Methods* **17**, 354–373 (2012).
85. Youthinmind. Development and Well-Being Assessment (DAWBA) https://dawba.info/ (2012).
86. Youthinmind. Strengths and Difficulties Questionnaire (SDQ) https://www.sdqinfo.org/a0.html (2022).
87. Goodman, R., Ford, T., Richards, H., Gatward, R. & Meltzer, H. The development and well-being assessment: description and initial validation of an integrated assessment of child and adolescent psychopathology. *J. Child Psychol. Psychiatry* **41**, 645–655 (2000).
88. Aebi, M. et al. The use of the development and well-being assessment (DAWBA) in clinical practice: a randomized trial. *Eur. Child Adolesc. Psychiatry* **21**, 559–567 (2012).
89. Kamper-DeMarco, K. E., Shankman, J., Fearey, E., Lawrence, H. R. & Schwartz-Mette, R. A. Linking social skills and adjustment. *Soc. Skil. Life Span* 47–66 (2020).
90. Chambers, J. M. lm function—stats package (version 3.6.2). https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/lm (1992).
91. Caldwell, A. R. Exploring equivalence testing with the updated TOSTER R package (preprint). *Open Science Framework* https://doi.org/10.31234/osf.io/ty8de (2022).
92. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300 (1995).
93. Judd, C. M., McClelland, G. H., & Ryan, C. S. *Data Analysis: a Model Comparison Approach to Regression, ANOVA, and Beyond* (Routledge, 2017).
94. Fassi et al. Registered Report: social media use in adolescents with and without mental health conditions. *Open Science Framework*, https://osf.io/s2dwu/?view_only=5acced2ddb884194 81d439a774614dd1 (2025).

95. Wysocki, A. C., Lawson, K. M. & Rhemtulla, M. Statistical control requires causal justification. *Adv. Methods Pract. Psychol. Sci.* **5**, 25152459221095823 (2022).
96. Shapiro, J. R., Klein, S. L. & Morgan, R. Stop 'controlling' for sex and gender in global health research. *BMJ Glob. Health* **6**, e005714 (2021).
97. Lakens, D. Equivalence tests: a practical primer for t tests, correlations, and meta-analyses. *Soc. Psychol. Personal. Sci.* **8**, 355–362 (2017).
98. Crone, D., Smith, A. & Gough, B. 'I feel totally at one, totally alive and totally happy': a psycho-social explanation of the physical activity and mental health relationship. *Health Educ. Res* **20**, 600–611 (2005).
99. Knowlden, A. P., Sharma, M. & Bernard, A. L. A theory of planned behavior research model for predicting the sleep intentions and behaviors of undergraduate college students. *J. Prim. Prev.* **33**, 19–31 (2012).
100. Rebar, A. L. & Taylor, A. Physical activity and mental health; it is more than just a prescription. *Ment. Health Phys. Act.* **13**, 77–82 (2017).
101. Mollayeva, T. et al. The Pittsburgh sleep quality index as a screening tool for sleep dysfunction in clinical and non-clinical samples: a systematic review and meta-analysis. *Sleep Med. Rev.* **25**, 52–73 (2016).
102. Ruscio, J. A probability-based measure of effect size: robustness to base rates and other factors. *Psychol. Methods* **13**, 19–30 (2008).
103. Mirzaei, A., Carter, S. R., Patanwala, A. E. & Schneider, C. R. Missing data in surveys: key concepts, approaches, and applications. *Res. Soc. Adm. Pharm.* **18**, 2308–2316 (2022).
104. Lakens, D., Scheel, A. M. & Isager, P. M. Equivalence testing for psychological research: a tutorial. *Adv. Meth. Pract. Psychol. Sci.* **1**, 259–269 (2018).
105. Champely, S. pwr package (version 1.3-0). https://cran.r-project.org/web/packages/pwr/pwr.pdf (2020).
106. Hamilton, J. L. et al. Social media use and prospective suicidal thoughts and behaviors among adolescents at high risk for suicide. *Suicide Life Threat. Behav.* **51**, 1203–1212 (2021).
107. Riehm, K. E. et al. Associations between time spent using social media and internalizing and externalizing problems among US youth. *JAMA Psychiatry* **76**, 1266–1273 (2019).
108. Bayer, J. B., Anderson, I. A. & Tokunaga, R. S. Building and breaking social media habits. *Curr. Opin. Psychol.* **45**, 101303 (2022).
109. Winds, K., Aebi, M. & Plattner, B. Problematic internet use among adolescent male and female psychiatric inpatients: a gender perspective. *Child Psychiatry Hum. Dev.* https://doi.org/10.1007/s10578-022-01408-6 (2022).
110. Clerkin, E. M., Smith, A. R. & Hames, J. L. The interpersonal effects of Facebook reassurance seeking. *J. Affect. Disord.* **151**, 525–530 (2013).
111. Hames, J. L., Hagan, C. R. & Joiner, T. Interpersonal Processes in Depression. *Annu. Rev. Clin. Psychol.* **9**, 355–377 (2013).
112. Elhai, J. D., Gallinari, E. F., Rozgonjuk, D. & Yang, H. Depression, anxiety and fear of missing out as correlates of social, non-social and problematic smartphone use. *Addict. Behav.* **105**, 106335 (2020).
113. Gotlib, I. H. Perception and recall of interpersonal feedback: negative bias in depression. *Cogn. Ther. Res.* **7**, 399–412 (1983).
114. Viner, R. M. et al. Adolescence and the social determinants of health. *Lancet* **379**, 1641–1652 (2012).
115. Sprecher, S. & Hendrick, S. S. Self-disclosure in intimate relationships: associations with individual and relationship characteristics over time. *J. Soc. Clin. Psychol.* **23**, 857–877 (2004).
116. Reece, A. G. & Danforth, C. M. Instagram photos reveal predictive markers of depression. *EPJ Data Sci.* **6**, 15 (2017).
117. Omarzu, J. A disclosure decision model: determining how and when individuals will self-disclose. *Personal. Soc. Psychol. Rev.* **4**, 174–185 (2000).
118. van den Bos, E., de Rooij, M., Miers, A. C., Bokhorst, C. L. & Westenberg, P. M. Adolescents' increasing stress response to social evaluation: pubertal effects on cortisol and alpha-amylase during public speaking. *Child Dev.* **85**, 220–236 (2014).
119. Vijayakumar, N. et al. Getting to know me better: an fMRI study of intimate and superficial self-disclosure to friends during adolescence. *J. Pers. Soc. Psychol.* **118**, 885–899 (2020).

Acknowledgements

The UK Medical Research Council DTP PhD programme (RG86932) funded L.F. The Jacobs Foundation (CERES SUAI/084 G114119), the UK Medical Research Council (MC_UU_00030/13) and a UKRI Future Leaders Fellowship (MR/X034925/1) funded A.M.F., L.F. and A.O. The Huo Family Foundation, and the ESRC (ES/Y010736/1 and ES/T008709/1) funded A.K.P. Both A.K.P. and A.O. were supported by the Economic and Social Research Council (ES/T008709/1). T.J.F. is supported by the National Institute for Health and Care Research (NIHR204413, NIHR153625 and NIHR202025), the Swedish Research Council for Health Working Life and Welfare (2022-01002_Forte), and the Medical Research Council (MC_PC_20052). All research at the Department of Psychiatry in the University of Cambridge is supported by the NIHR Cambridge Biomedical Research Centre (NIHR203312) and the NIHR Applied Research Collaboration East of England. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care. T.J.F.'s research group receives funding from Place2Be, a third sector organisation that provides mental health training and intervention to UK schools. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. The MHCYP 2017 survey was funded by the Department of Health and Social Care, commissioned by NHS Digital, and carried out by the National Centre for Social Research, the Office for National Statistics and Youthinmind. We are very grateful to all the adolescents and families who took part in the study, the personnel for their help in recruiting them and the whole NHS Digital team that includes interviewers, technicians, research scientists, volunteers, managers, receptionists, nurses and the clinical raters. A special thank you to D. Lakens, M. Vuorre and A.R. Caldwell for their valuable advice on the statistical analyses.

Author contributions

We present author contributions according to the CRediT (Contributor Roles Taxonomy). L.F.: conceptualization, methodology, formal analysis, writing— original draft, writing—review and editing. A.M.F.: conceptualization, methodology, formal analysis, writing—review and editing. A.K.P.: methodology, writing—review and editing, supervision. T.J.F.: writing—review and editing, supervision, resources, project administration. A.O.: conceptualization, methodology, writing— original draft, writing—review and editing, supervision. Of note, none of the authors that conceptualized the analysis previously accessed the MHCYP dataset. For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41562-025-02134-4.

Correspondence and requests for materials should be addressed to Luisa Fassi.

Peer review information *Nature Human Behaviour* thanks Candice Biernesser, Johannes Breuer, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Protocol registration The Stage 1 protocol for this Registered Report was accepted in principle on 7 December 2023. The protocol, as accepted by the journal, is available via Figshare at https://figshare.com/s/730e3b0d4da83e9b6a46.

nature portfolio

Corresponding author(s): Fassi Luisa
Last updated by author(s): Jan 27, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	This study uses secondary data, details on how to access the data are reported in the Data Availability Statement. Specifically, the MHCYP dataset is held on behalf of NHS Digital by the UK Data Service. Restrictions apply to the availability of this data for privacy and ethical reasons, which were used under license for this study. Data access can be requested by applying to the Data Access Request Service (DARS; number for this study: DARS-NIC-424336-T7K7T-v0.6). Researchers interested in accessing the data can find further information on the DARS website (see https://digital.nhs.uk/services/data-access-request-service-dars/dars-guidance).
Data analysis	The analysis code used to run the power analysis (stage 1) and all other analyses (stage 2) can be found at: https://osf.io/s2dwu/?view_only=5accd2ddb8849481d439a7746f4dd1. Details on each code script are reported in the READ ME file. We used R version 4.4.0 (2024-04-24), and the following packages: Cairo (version: 1.6-2), readstata13 (version: 0.10.1), dplyr (version 1.1.2), TOSTER (version: 0.8.0), ggplot2 (version 3.5.1), patchwork (version: 1.2.0.9000), tidyverse (version: 2.0.0).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The MHCYP dataset is held on behalf of NHS Digital by the UK Data Service. Restrictions apply to the availability of this data for privacy and ethical reasons, which were used under license for this study. Data access can be requested by applying to the Data Access Request Service (DARS; number: DARS-NIC-424336-T7K7T-v0.6). Researchers interested in accessing the data can find further information on the DARS website (i.e., https://digital.nhs.uk/services/data-access-request-service-dars/dars-guidance).

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Our final sample included 3,340 young people aged 11 to 19 (age: mean = 14.77, sd = 2.48). The sample was 50% male and 50% female. Both sex and age were measured using self-report. We provide descriptives for age and sex separately for each group in Table 3 and the section 'population characteristics' below. We also provide separate descriptive statistics separately for males and females (sex variable) for each social media item in the Supplementary Results (Supplementary Table 5-8). We do not provide separate descriptive statistics for age as this question falls outside the scope of this study.
Reporting on race, ethnicity, or other socially relevant groupings	While it is common in research to use statistical control to remove confounding effects from a regression coefficient, appropriate control variables should be identified only after justifying a causal structure that includes the outcome, exposure and all theorised confounders. When the selected control variables are inappropriate or remain unjustified, controlling can result in biased regression estimates. Further, recent literature warns against controlling for demographic factors such as sex without thought and instead prompts researchers to interrogate how this variable intersects with the exposure and outcomes under investigation. In the present work, treating sex or age as a confounding variable would mean ignoring the possibility that there are meaningful sex or age differences in the examined relationships. As our goal is to investigate the association between social media use and mental health diagnosis, we provided a descriptive account of the age and sex of adolescents in each group. As shown below, age and sex of adolescents with vs without a mental health condition are reported for descriptive purposes.
Population characteristics	The age and sex of each examined group are reported in Table 3 in the main manuscript and below. - No mental health (N = 2821, age mean = 14.71, male proportion: 0.50) - Any mental health condition (N = 519, age mean = 15.10, male proportion = 0.47) - Externalising condition(N = 104, age mean = 14.27, male proportion = 0.72) - Internalising condition (N = 282, age mean = 15.94, male proportion = 0.80) - Other conditions (N = 76, age mean = 14, male proportion = 0.80) - Comorbidity between internalising and externalising condition (N = 57, age mean = 13.93, male proportion = 0.51).
Recruitment	The MHCYP study is one of a series of national surveys on the mental health of children and young people in England administered in 1999, 2004, 2017, 2021 and 2022. In this Registered Report, we analysed the 2017 wave collected between January and October 2017: the most recent wave to be made available to researchers as well as the first wave to collect comprehensive data on adolescents' social media use and to include 17 to 19-year-olds. We only analysed data from adolescents who reported being social media users aged 11-19 years, a total of 3,340 participants (50 % male, 50% female) out of the full sample of 9,117. The survey was collected using a stratified probability sample of children and young people living in England who were registered with a general practitioner. Data was collected via face-to-face interviews with adolescents and their parents. At the same time, if the family agreed, questionnaires were mailed to teachers (for the available data and key demographics see MHCYP 2017: https://digital.nhs.uk/data-and-information/publications/statistical/mental-health-of-children-and-young-people-in-england/2017/2017).
Ethics oversight	The MHCYP 2017 survey was reviewed and approved by the West London & GTAC Research Ethics Committee (REC reference: 16/LO/0155) and the Health Research Authority Confidentiality Advisory Group (CAG reference: 16/CAG/0016) in 2016. Both parents and children provided consent to take part in data collection and were compensated with a 10£ voucher for their time. Parents of children under 16 years were interviewed first, and permission was sought to interview their child afterwards; the child then provided assent. Conversely, 17–19-year-olds were directly asked for their consent, with permission subsequently sought for their parents to be interviewed.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description

This study analyses data from a quantitative cross-sectional survey collected in 2017 (Mental Health of Children and Young People).

Research sample

For the 2017 survey, a stratified, multistage random probability sample of 18,029 children was drawn from the NHS Patient Register in October 2016. Children and young people were eligible to participate if they were aged 2 to 19, lived in England, and were registered with a GP. The sample was designed to be representative of the population of children and young people aged 2 to 19 living in England. The final sample consisted of 9,117 children and young people. For this study, we analysed data from 3,340 participants. This subset was selected based on age (focusing on adolescents aged 11–19 years) and social media use (including only those who reported being social media users).

Sampling strategy

A stratified multistage random probability sample was used for the survey, involving a two-stage process. Full information on the sampling can be found at: https://files.digital.nhs.uk/9D/1CB03A/MHCYP%202017%20Survey%20Design%20and%20Methods.pdf

To determine the sample size needed to answer our research questions, we run power analyses and power sensitivity analyses. A priori power calculations were conducted and reported in the Stage 1 registered report. For these calculations, we set were the smallest effect size of interest (SESOI, d = 0.4; see Supplementary Methods for details), an alpha level of 0.05, and an estimated sample size of N = 3,854. The code for these calculations is available on the Open Science Framework (OSF) at https://osf.io/s2dwu/?view_only=5acced2ddb884f9481d439a7746f4dd1. Results showed that we were sufficiently powered (> 95%) to detect our smallest effect size of interest for all research questions.

In addition to the a priori power calculations, we conducted power sensitivity analyses using the final sample size, which was known only after starting data analysis. In this case we determined the smallest effect size that could be detected with 95% power, given an alpha level of 0.0125 (corrected for multiple comparisons) and the final sample size (N = 3340). As reported in Supplementary Table 18, these analyses demonstrated power to detect our smallest effect size of interest (d = 0.4) across our research questions.

Data collection

Data were collected as part of a national survey. Researchers were therefore blind to the study conditions and hypotheses. All interviews were conducted individually, involving only the clinical rater and the interviewee (child, parent, or teacher) and were carried out either in person or in a computerized format.

For participants aged 11 to 16 years, the process began with an initial interview with the parent or legal guardian, followed by a separate interview with the child. Young people aged 17 to 19 were interviewed directly, with their parent also interviewed if both parties consented. Teachers of 5 to 16-year-olds were invited to complete an online or paper questionnaire if consent was provided.

Mental health assessments were conducted using the detailed and comprehensive Development and Well-Being Assessment (DAWBA; Goodman et al., 2000), which evaluates a range of mental health conditions, including emotional, hyperactivity, and behavioral disorders, as well as less common conditions like autism. After completing the interviews, trained clinical raters reviewed the data to assess the presence of mental health conditions for each participant.

Timing

Data collection occurred over nine months, from January to September 2017. For participants aged 11–16 years, data collection was conducted between January and June 2017 to maintain consistency with previous surveys in the series and to ensure that teacher questionnaires could be completed and returned before the end of the school summer term. For children aged 16 or younger, data collection began with an interview with the parent. Parental permission was then sought to interview the child. If consent was given, the child participated in an interview, which included a self-completion section for sensitive questions. For young people aged 17–19 years, agreement to participate was obtained directly from the individual. Further details on data collection can be found at: https://files.digital.nhs.uk/9D/1CB03A/MHCYP%202017%20Survey%20Design%20and%20Methods.pdf

Data exclusions

In this study, we excluded participants aged 2–10 years (N= 5060) and those who reported not using social media (N = 717). Both these exclusions were pre-established (in our Stage 1 registered report) for two reasons: 1) children younger than 11 are in a different developmental time window (childhood/pre-adolescence), 2) we were interested in young people that used social media to answer our research questions. Each social media use item was analyzed independently, allowing us to retain data for individual items even when responses were incomplete for other items. For example, if a child answered the question about time spent on social media but not the question about online social comparison, their response to the time spent question was still included in the analysis. Consequently, the exact sample size varies for each question and is reported in Supplementary Tables 5–8.

Non-participation

No participant declined participation.

Randomization

We did not perform randomization nor we controlled for third variables. While it is common in research to use statistical control to remove confounding effects from a regression coefficient, appropriate control variables should be identified only after justifying a causal structure that includes the outcome, predictors, and all theorised confounders. When the selected control variables are inappropriate or remain unjustified, controlling can result in biased regression estimates. Further, recent literature warns against controlling for demographic factors such as sex without thought and instead prompts researchers to interrogate how this variable intersects with the predictors and outcomes under investigation. In the present work, treating sex or age as a confounding variable would mean ignoring the possibility that there are meaningful sex or age differences in the examined relationships. As our goal is to investigate the overall association between social media use and mental health diagnosis, we provided a descriptive account of the age and sex of adolescents included in each tested model rather than control for these demographics.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks	NA
Novel plant genotypes	NA
Authentication	NA

JAMA Pediatrics | Original Investigation | ADOLESCENT MENTAL HEALTH

Social Media Use and Internalizing Symptoms in Clinical and Community Adolescent Samples
A Systematic Review and Meta-Analysis

Luisa Fassi, BSc, MSc; Kirsten Thomas, BSc, MRes; Douglas A. Parry, BA, MA, PhD; Amelia Leyland-Craggs, BA; Tamsin J. Ford, MD, PhD; Amy Orben, MA, DPhil

Supplemental content

IMPORTANCE In response to widespread concerns about social media's influence on adolescent mental health, most research has studied adolescents from the general population, overlooking clinical groups.

OBJECTIVE To synthesize, quantify, and compare evidence on the association between social media use and internalizing symptoms in adolescent clinical and community samples.

DATA SOURCES Peer-reviewed publications from MEDLINE, Web of Science, PsycInfo, and Scopus (initially reviewed in May 2022 and updated in October 2023) and pr
from Europe PubMed Central (February 2023) published in English between 20

STUDY SELECTION Two blinded reviewers initially identified 14 211 cross-sectional
longitudinal studies quantifying the association between social media use and in
symptoms, excluding experimental studies and randomized clinical trials.

DATA EXTRACTION AND SYNTHESIS PRISMA and MOOSE guidelines were followe
pooling data using a random-effects model and robust variance estimation. The q
of evidence was assessed using the Quality of Survey Studies in Psychology Chec

MAIN OUTCOMES AND MEASURES Articles were included if they reported at least
1 quantitative measure of social media use (time spent, active vs passive use, acti
content, user perception, and other) and internalizing symptoms (anxiety, depres
or both).

RESULTS The 143 studies reviewed included 1 094 890 adolescents and 886 effec
11% of which examined clinical samples. In these samples, a positive and significan
meta-correlation was found between social media use and internalizing symptoms
both for time spent (n = 2893; r, 0.08; 95% CI, 0.01 to 0.15; P = .03; I^2, 57.83) and
engagement (n = 859; r, 0.12; 95% CI, 0.09 to 0.15; P = .002; I^2, 82.67). These asso
mirrored those in community samples.

CONCLUSIONS AND RELEVANCE The findings in this study highlight a lack of research on clinical populations, a critical gap considering public concerns about the increase in adolescent mental health symptoms at clinical levels. This paucity of evidence not only restricts the generalizability of existing research but also hinders our ability to evaluate and compare the link between social media use and mental health in clinical vs nonclinical populations.

Clinical

Author Affiliations: Medical Research Council Cognition and Brain Sciences Unit, University of Cambridge, Cambridge, United Kingdom (Fassi, Leyland-Craggs, Orben); Department of Psychiatry, University of Cambridge, Cambridge, United Kingdom (Fassi, Ford); Department of Psychology, University of Cambridge, Cambridge, United Kingdom (Thomas); Department of Information Science, Stellenbosch University, Stellenbosch, South Africa (Parry).

Corresponding Author: Luisa Fassi, BSc, MSc, Medical Research Council Cognition and Brain Sciences Unit, University of Cambridge, Cambridge, United Kingdom, 15 Chaucer Rd, Cambridge CB2 7EF, United Kingdom (luisa.fassi@mrc-cbu.cam.ac.uk).

JAMA Pediatr. 2024;178(8):814-822. doi:10.1001/jamapediatrics.2024.2078
Published online June 24, 2024.

JAMA Pediatrics | Original Investigation | ADOLESCENT MENTAL HEALTH

Social Media Use and Internalizing Symptoms in Clinical and Community Adolescent Samples

A Systematic Review and Meta-Analysis

Luisa Fassi, BSc, MSc; Kirsten Thomas, BSc, MRes; Douglas A. Parry, BA, MA, PhD; Amelia Leyland-Craggs, BA; Tamsin J. Ford, MD, PhD; Amy Orben, MA, DPhil

Supplemental content

IMPORTANCE In response to widespread concerns about social media's influence on adolescent mental health, most research has studied adolescents from the general population, overlooking clinical groups.

OBJECTIVE To synthesize, quantify, and compare evidence on the association between social media use and internalizing symptoms in adolescent clinical and community samples.

DATA SOURCES Peer-reviewed publications from MEDLINE, Web of Science, PsycInfo, and Scopus (initially reviewed in May 2022 and updated in October 2023) and preprints from Europe PubMed Central (February 2023) published in English between 2007 and 2023.

STUDY SELECTION Two blinded reviewers initially identified 14 211 cross-sectional and longitudinal studies quantifying the association between social media use and internalizing symptoms, excluding experimental studies and randomized clinical trials.

DATA EXTRACTION AND SYNTHESIS PRISMA and MOOSE guidelines were followed, pooling data using a random-effects model and robust variance estimation. The quality of evidence was assessed using the Quality of Survey Studies in Psychology Checklist.

MAIN OUTCOMES AND MEASURES Articles were included if they reported at least 1 quantitative measure of social media use (time spent, active vs passive use, activity, content, user perception, and other) and internalizing symptoms (anxiety, depression, or both).

RESULTS The 143 studies reviewed included 1 094 890 adolescents and 886 effect sizes, 11% of which examined clinical samples. In these samples, a positive and significant meta-correlation was found between social media use and internalizing symptoms, both for time spent ($n = 2893$; r, 0.08; 95% CI, 0.01 to 0.15; $P = .03$; I^2, 57.83) and user engagement ($n = 859$; r, 0.12; 95% CI, 0.09 to 0.15; $P = .002$; I^2, 82.67). These associations mirrored those in community samples.

CONCLUSIONS AND RELEVANCE The findings in this study highlight a lack of research on clinical populations, a critical gap considering public concerns about the increase in adolescent mental health symptoms at clinical levels. This paucity of evidence not only restricts the generalizability of existing research but also hinders our ability to evaluate and compare the link between social media use and mental health in clinical vs nonclinical populations.

JAMA Pediatr. 2024;178(8):814-822. doi:10.1001/jamapediatrics.2024.2078
Published online June 24, 2024.

Author Affiliations: Medical Research Council Cognition and Brain Sciences Unit, University of Cambridge, Cambridge, United Kingdom (Fassi, Leyland-Craggs, Orben); Department of Psychiatry, University of Cambridge, Cambridge, United Kingdom (Fassi, Ford); Department of Psychology, University of Cambridge, Cambridge, United Kingdom (Thomas); Department of Information Science, Stellenbosch University, Stellenbosch, South Africa (Parry).

Corresponding Author: Luisa Fassi, BSc, MSc, Medical Research Council Cognition and Brain Sciences Unit, University of Cambridge, Cambridge, United Kingdom, 15 Chaucer Rd, Cambridge CB2 7EB, United Kingdom (luisa.fassi@mrc-cbu.cam.ac.uk).

Adolescent mental health has declined substantially in recent years. The proportion of UK adolescents (aged 10-24 years)[1] with a probable mental health condition increased from 10% to 25% between 2017 and 2022.[2,3] Globally, 1 in 5 children and adolescents have a mental health condition, most commonly internalizing disorders (eg, anxiety or depression).[4] The impact of such conditions is wide reaching and long lasting, affecting school attendance, interpersonal relationships, employment prospects, physical health, and suicide risk, with suicide now constituting the second-leading cause of death among 15- to 29-year-olds worldwide.[4] Many raise concerns that social media, now ubiquitous (97% of young people are daily users),[5] is accelerating current mental health declines.[6,7]

Scientific research investigating social media's impact on adolescent mental health has failed to provide clarity. There is converging evidence for a small negative cross-sectional association between time spent on social media and well-being.[8-11] However, longitudinal studies and those measuring social media use beyond time spent or mental health beyond general well-being show diverging results.[12-14]

To understand this heterogeneity, researchers have studied whether individual differences (eg, age, sex, or ethnicity) might moderate the relationship between social media use and mental health.[15-18] However, the potential impact of the mental health status of the examined sample has been largely overlooked. Studies routinely recruit adolescents from the general population through schools, universities, or nationally representative surveys.[12-14] While these samples can include individuals experiencing mental health symptoms at clinical levels, they often fail to distinguish them from those experiencing symptoms at subclinical or nonclinical levels.

Individuals with mental health conditions face unique challenges, such as interpersonal or sleep difficulties and educational disruptions.[19] Adolescents with internalizing conditions, in particular, exhibit heightened sensitivity to social comparison and fear of negative evaluation.[20,21] They might therefore use or be impacted by social media differently compared to peers. Failure to account for the nature and severity of mental health indicators could therefore restrict our ability to draw accurate inferences about social media's relationship with mental health.

We addressed the extent and impact of this oversight in 3 steps. First, we completed a preregistered systematic review to quantify the proportion of studies investigating social media use and internalizing symptoms in adolescent clinical samples compared to community or nonclinical samples. Second, we performed a meta-analysis to extract the pooled association between social media use and internalizing symptoms in clinical samples, differentiating between time spent and other measures of social media engagement. Third, we compared the strength and direction of this association across clinical and community samples, testing whether sample type was a moderator.

This work allowed us to gauge whether and how current research in this area of substantial scientific and public interest can be used to make clinically informative recommendations. It also complements preexisting qualitative reviews,[22] synthesizing the quantitative effect sizes in clinical populations and comparing these with community samples. Together, these findings can inform academics by identifying gaps for future research; clinicians, by summarizing research studying relevant populations; and policymakers, by guiding evidence-based decision-making for adolescents at risk.

Key Points

Question What is the proportion of research on the association between social media use and mental health in adolescent clinical populations and does it differ between clinical and community samples?

Findings In this systematic review and meta-analysis of 143 studies, few focused on clinical populations, and these showed a positive association between social media use and internalizing symptoms. These results mirrored findings from community samples.

Meaning The paucity of research on clinical populations limits the generalizability of existing research and hinders a comprehensive evaluation of the association between social media and mental health.

Methods

Search Strategy, Selection, and Extraction

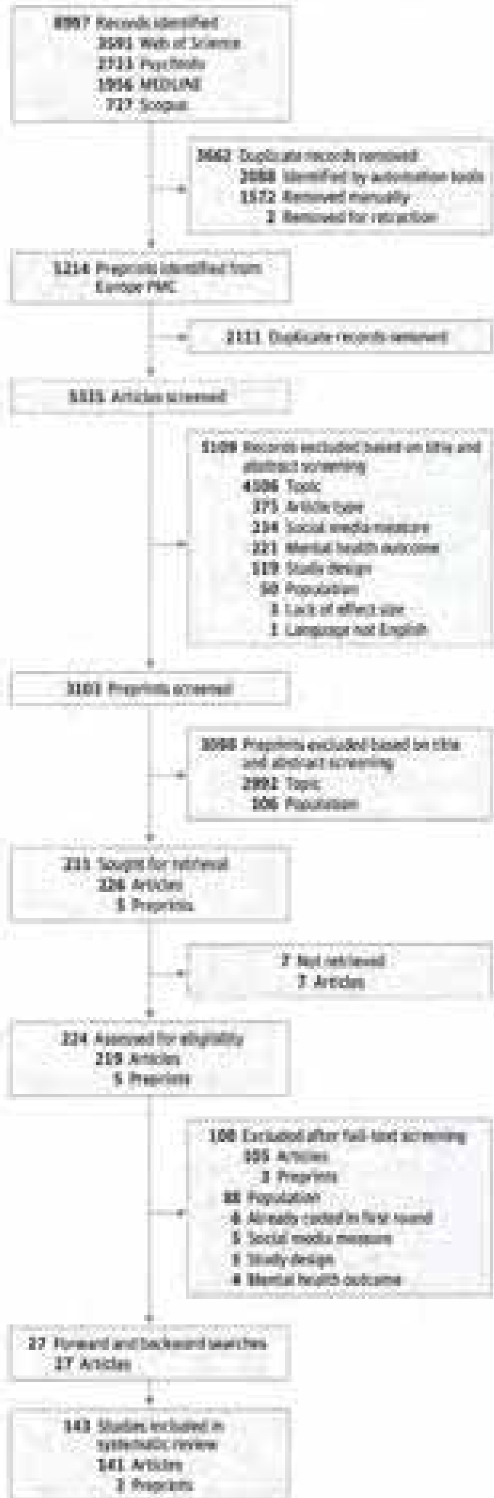
The protocol for this study was preregistered with Prospero (CRD42022321473), following the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) reporting guideline. MEDLINE, Web of Science, PsycInfo, and Scopus were searched (eAppendix 1 in Supplement 1) initially in May 2022 and updated in October 2023; forward/backward citation tracing via Google Scholar and preprint search via Europe PubMed Central in February 2023. We identified 14 211 records (8997 articles and 5214 preprints) and 8438 (5335 articles and 3103 preprints) remained after duplicate removal. Considering the nature of the study design, no ethical review was needed.

Selection criteria (eAppendix 2 in Supplement 1) were peer-reviewed English-language articles and preprints published in or after January 2007; quantitative time- or engagement-based social media use measures, self-reported or logged; quantitative symptom-based or other validated questionnaires of anxiety, depression, or both; and adolescent populations aged 10 to 24 years (if not provided: mean age ±1 SD in age range).

In terms of social media, we categorized types of engagement into 6 preregistered categories to allow meaningful description and pooling of studies (eAppendix 2 in Supplement 1): time spent and frequency, activities (eg, messaging and posting), content (eg, exposure to appearance-related content), user perception (eg, impact of likes on mood), active vs passive use, and other.

We categorized study samples into clinical, community or nonclinical. Clinical samples included adolescents who either scored above the clinical threshold on a validated questionnaire, reporting an active diagnosis, accessed mental health services, or were psychiatrically hospitalized. Community samples

Figure 1. Flow Diagram



Reporting of study identification, screening and inclusion for the systematic review. PMC indicates PubMed Central.

included adolescents across the entire distribution of internalizing symptoms without separation into clinical and nonclinical levels, while nonclinical samples excluded adolescents in the clinical range. We restricted our focus to studies that examined internalizing symptoms, excluding other conditions (eg, externalizing or neurodevelopmental) unless they were comorbid with internalizing symptoms.[23]

Following title and abstract screening by 2 independent reviewers, 231 records (226 articles and 5 preprints) remained and were full-text screened (Figure 1; eAppendix 3 in Supplement 1). Three independent reviewers double coded 10% of studies to harmonize the coding strategy (eAppendix 4 in Supplement 1; reliability: 95%) and then extracted study information, samples, measures, methodologies, and effect sizes. Risk of bias and quality of studies was assessed using an adapted version of the Quality of Survey Studies in Psychology (eAppendix 5 in Supplement 1).[24]

Statistical Analysis

We completed all analyses in R version 4.1.2 (R Foundation; full list of packages on OSF).[25] We first conducted descriptive analyses of the studies included in the systematic review. Specifically, we calculated the number of studies and effect sizes (with associated percentages), split by sample type, mental health measure, social media measure, social media data collection, study design, and global population (countries were coded based on the International Telecommunications Union classification[26]).

Next, we conducted meta-analyses to test the pooled association between social media use and internalizing symptoms for clinical and community samples. These meta-analyses were restricted to cross-sectional studies and the initial cross-sectional wave of longitudinal studies (see eAppendix 2 in Supplement 1 for details on this choice). Our confirmatory meta-analyses were restricted to studies measuring time spent on social media, to allow meaningful pooling of effect sizes due to measurement similarity, while in the exploratory meta-analyses, we examined measures of social media engagement.

The association was defined as positive when increased social media use was associated with increased internalizing symptoms. We used an a priori statistical significance level of $\alpha = .05$ and interpreted effect sizes in line with Cohen (1988; small: $r < 0.10$, medium: $r = 0.30$, large: $r > 0.50$). For studies reporting effect sizes other than a correlation coefficient, we performed transformations where possible (eAppendix 6 in Supplement 1). We transformed all correlations from Fisher z back to Pearson r for reporting.

We used a random-effects model to calculate summary effect sizes due to the high level of heterogeneity. To account for variance inflation emerging from dependent observations for measures collected from the same participants, we used cluster-robust variance estimation based on the sandwich method with adjusted estimators for small samples and the correlated effects weighting scheme using robumeta ($r = 0.80$ for the within-study effect size correlation).[27-29] Sensitivity analyses showed that using different r values did not affect the inferences made.[25]

Given that longitudinal studies have multiple waves per participant, the meta-analysis included only the effect size from the first wave to minimize variance inflation. However, no differences emerged in the strength and direction when including all waves (eAppendix 7 in Supplement 1).

Risk of Bias Assessment and Moderation

To assess potential bias due to small study effects, including publication bias, we visually inspected funnel plot symmetry and performed the Egger regression test.[30,31] Further, we used a contour-enhanced funnel plot with superimposed areas of statistical significance (corresponding to P = .10, .05, and .01), interpreting an overrepresentation of effect sizes in the highlighted areas as indicative of publication bias.[31] We conducted influence diagnostics (ie, the Cook distance, covariance ratio, and diagonal elements of the hat matrix) using metafor[32] to identify outliers and performed leave-one-out sensitivity analyses with such outliers removed (eAppendix 8 in Supplement 1). To examine heterogeneity in effect sizes, we computed I^2, interpreting values around 25%, 50%, and 75% to indicate low, moderate, and high heterogeneity, respectively.

We conducted 3 preregistered moderator analyses to investigate factors contributing to heterogeneity, namely, sample type (clinical or community samples; nonclinical samples were excluded due to a lack of power), mental health measure (anxiety, depression, or internalizing symptoms) and COVID-19[33-35] (before vs during). We classified studies as happening during the COVID-19 pandemic if any data collection was performed after January 2020.[4] Lastly, we conducted exploratory moderation analyses for age, sex, and the type of social media measure for the meta-analysis on social media engagement.

Results

Systematic Review: Quantifying the Proportion of Clinical Samples

After duplicate removal, we screened 8438 manuscripts (5335 articles and 3103 preprints), including 143 studies in the systematic review (141 articles and 2 preprints; Figure 1). Included studies had a combined sample size of 1 094 890 adolescents (mean, 7657; SD, 40 026; median, 680; minimum, 41; maximum, 388 275) and reported 886 effect sizes for the association between social media use and internalizing symptoms (eAppendix 9 in Supplement 1).

Studies investigating adolescent clinical samples were rare: 11% of effect sizes, corresponding to 99 effect sizes from 12 studies (**Figure 2A**; eAppendix 10 in Supplement 1). Most studies examined community samples (88% of effect sizes; 774 effect sizes from 133 studies), with very few focusing on nonclinical samples (1% of effect sizes; 13 effect sizes from 4 studies). The most common mental health measure was depression (67% of effect sizes; 595 effect sizes from 118 studies), while anxiety (26% of effect sizes; 228 effect sizes from 52 studies) and internalizing symptoms (7% of effect sizes; 63 effect sizes from 16 studies) were less frequently assessed (Figure 2B).

Regarding social media measures, 92% of effect sizes were derived from studies using self-reports (816 effect sizes from 138 studies), while 8% used logged measures (70 effect sizes from 8 studies). Nearly half of the effect sizes were extracted from studies measuring time spent (43%; 381 effect sizes from 91 studies). Less common engagement-based measures included user perception (18%; 160 effect sizes from 36 studies), activity (15%; 131 effect sizes from 31 studies), active vs passive use (7%; 65 effect sizes from 14 studies), content (3%; 29 effect sizes from 4 studies), and other metrics (14%; 120 effect sizes from 21 studies). Most studies (66%; 94 studies) were cross-sectional, while 34% (49 studies) were longitudinal (eAppendix 11 in Supplement 1). In line with previous work,[16] the most commonly studied populations were from the Global North (82%; 117 studies), compared to the Global South (18%; 26 studies).[29]

Overall, approximately half of the included studies (55%; 78 studies) were of acceptable quality based on the Quality of Survey Studies in Psychology Checklist. The remaining 45% (65 studies) were classified as being of questionable quality (eAppendix 5 in Supplement 1).

Meta-Analysis: Quantifying Associations in Clinical Samples

Social Media Time Spent

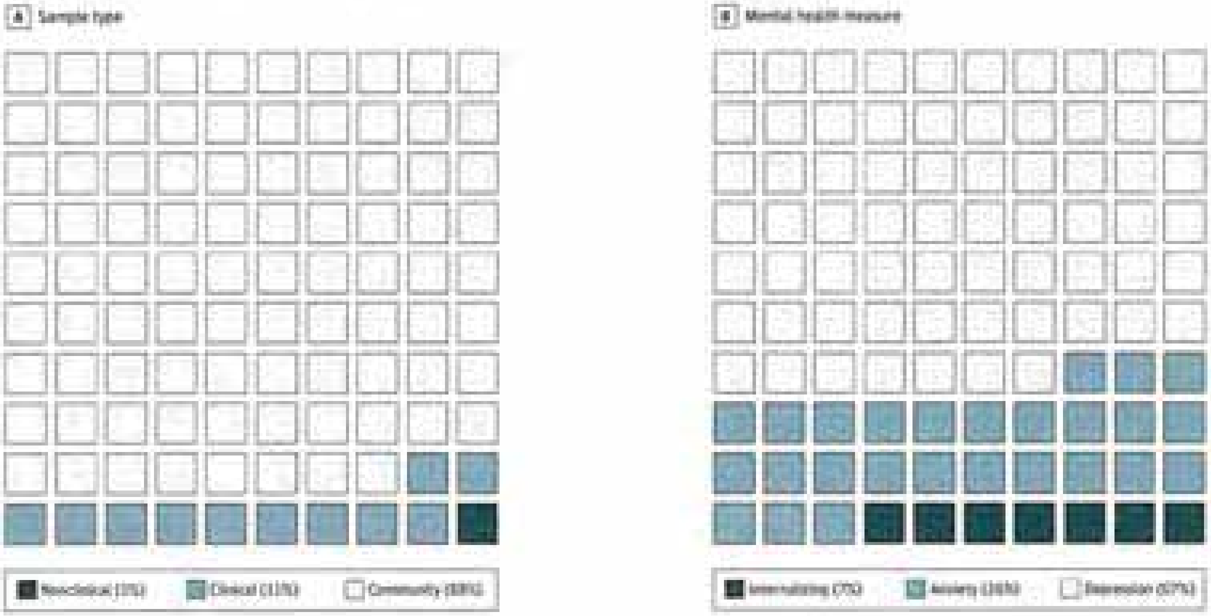
Seven studies of clinical populations (15 effect sizes) measured time spent on social media. The total sample size was 2893 (mean, 413; SD, 585; median, 224; minimum, 49; maximum, 1722). In our confirmatory meta-analysis, we found a positive and significant meta-correlation between time spent on social media and mental health symptoms (r, 0.08; 95% CI, 0.01 to 0.15; P = .03), with moderate heterogeneity (I^2, 57.83) (**Figure 3**[36-42]). Further, the Egger regression test showed no evidence of small study bias (β, −2.19; SE, 0.46; P = .98) (funnel plots in eAppendix 12 in Supplement 1).

Social Media Engagement

The need to move beyond time spent measures of social media use has been widely acknowledged, as these measures are simplistic and fail to distinguish between types of activities or content that can differentially relate to mental health.[43,44] Researchers have therefore advocated for using engagement-based measures, which we examined in an exploratory meta-analysis.

Four studies of clinical populations (19 effect sizes) used engagement-based measures (eAppendix 2 in Supplement 1), specifically social media activities (10 effect sizes; eg, selfie posting) and user perception (9 effect sizes; importance of social media use to daily life), with a total sample size of 859 (mean, 215; SD, 122; median, 233; minimum, 49; maximum, 343). We found a positive and significant meta-correlation between these social media measures and mental health symptoms (r, 0.12; 95% CI, 0.09 to 0.15; P = .002), with high heterogeneity (I^2, 82.67) (**Figure 4**[16,38,45]). Further, the Egger regression test showed no evidence of small study bias (β, −0.55; SE, 0.15; P = .93) (funnel plots in eAppendix 12 in Supplement 1).

Figure 2. Proportion of Included Effect Sizes by Sample Type and Mental Health Measure



Grid of 10 × 10 (100%) squares representing the percentage of literature in the systematic review by sample type and mental health measure. The presented proportion is calculated based on the total number of effect sizes (N = 898).

Meta-Analysis: Comparing Associations Between Clinical and Community Samples

Social Media Time Spent

We also ran a meta-analysis of the 49 studies (and 99 effect sizes) testing community samples (n = 479 215; mean, 9780; SD, 55 482; median, 442; minimum, 41; maximum, 388 275). We found a positive and significant meta-correlation between time spent on social media and internalizing symptoms (r, 0.12; 95% CI, 0.09 to 0.15; $P < .001$) (Figure 3; eAppendix 13 in Supplement 1). This is similar to the meta-correlation found in clinical samples (r, 0.08; 95% CI, 0.01 to 0.15; $P = .03$; I^2, 57.83) but shows higher levels of heterogeneity (I^2, 98.23).

To further test whether sample type influenced the association between time spent on social media and internalizing symptoms, we ran a combined meta-analysis (56 studies with 117 effect sizes; n = 482 273; mean, 8612; SD, 51 926; median, 372; minimum, 41; maximum, 388 275) and tested sample type as a moderator. We found an overall positive meta-correlation across all sample types (r, 0.12, 95% CI, 0.09 to 0.14; $P < .001$; I^2, 98.01) with no evidence of small study bias (β, −0.86; SE, 0.48; $P = .96$) (funnel plots in eAppendix 12 in Supplement 1).

After excluding nonclinical samples due to a lack of power (3 effect sizes from 3 studies), we tested sample type as a moderator (clinical vs community sample). We found nonsignificant results (β, 0.05; SE, 0.03; t, 1.6; 95% CI, −0.02 to 0.12; $P = .145$), and heterogeneity remained high (I^2, 98.0) (Table). Sample type was therefore not considered a key factor explaining differences in the association between time spent on social media and internalizing symptoms among adolescents.

We also tested whether the mental health measure (anxiety, depression, and internalizing symptoms) and COVID-19 (before vs during) were moderators. Neither the mental health measure (depression vs internalizing: β, −0.07; SE, 0.08; t, −0.81; 95% CI, −0.30 to 0.17; $P = .47$; anxiety vs internalizing: β, −0.07; SE, 0.08; t, −0.81; 95% CI, −0.27 to 0.14; $P = .45$) nor COVID-19 (β, 0.04; SE, 0.04; t, 1.16; 95% CI, −0.04 to 0.12; $P = .27$) explained heterogeneity in the meta-correlation between time spent on social media and mental health (Table). There was also no moderation for age or sex (eAppendix 14 in Supplement 1).

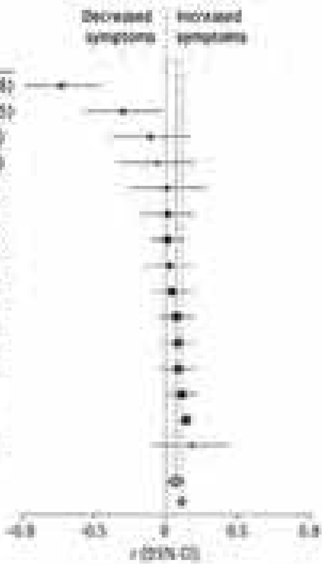
Social Media Engagement

We repeated the same analyses for studies measuring social media engagement. As with the meta-correlation found in clinical samples (r, 0.12; 95% CI, 0.09 to 0.15; $P = .002$; I^2, 82.67), we found a positive and significant association between social media engagement and mental health symptoms (r, 0.14; 95% CI, 0.10 to 0.18; $P < .001$) (Figure 4) in community samples (217 effect sizes from 62 studies; n = 65,799; mean, 1061; SD, 1607; median, 546; minimum, 41; maximum, 10 563). There were high levels of heterogeneity (I^2 = 94.85) (eAppendix 13 in Supplement 1).

We included 65 studies with 236 effect sizes in our combined meta-analysis. Across all sample types (n = 68 807; mean, 1,058; SD, 1605; median, 551; minimum, 41; maximum, 10 563), there was a positive meta-correlation between social media engagement and internalizing symptoms (r, 0.14; 95% CI, 0.10 to 0.17; $P < .001$) with high heterogeneity (I^2, 94.63). There was no evidence of small study bias (β, 0.54; SE, 0.67; $P = .22$) also confirmed by

Figure 3. Time Spent on Social Media and Internalizing Symptoms in Clinical Groups

Source	Sample, No.	Platform	Mental health	r (95% CI)
Fardouly et al,[37] 2020	49	Snapchat	Depression	−0.67 (−0.80 to −0.48)
Fardouly et al,[37] 2020	49	Instagram	Anxiety	−0.32 (−0.55 to −0.03)
Fardouly et al,[37] 2020	49	YouTube	Anxiety	−0.11 (−0.38 to 0.17)
Fardouly et al,[37] 2020	49	Snapchat	Anxiety	−0.07 (−0.33 to 0.21)
Fardouly et al,[37] 2020	49	Instagram	Depression	0.01 (−0.28 to 0.29)
Hamilton et al,[38] 2021	100	Any	Anxiety	0.01 (−0.19 to 0.21)
Nesi et al,[33] 2021	243	Any	Internalizing	0.01 (−0.12 to 0.14)
Hamilton et al,[38] 2021	100	Any	Depression	0.02 (−0.18 to 0.22)
Ucar et al,[39] 2018	224	Any	Anxiety	0.04 (−0.09 to 0.17)
Nesi et al,[33] 2021	243	Any	Internalizing	0.08 (−0.05 to 0.20)
Noel et al,[40] 2022	212	Any	Depression	0.09 (−0.05 to 0.22)
Ucar et al,[39] 2018	224	Any	Depression	0.09 (−0.04 to 0.22)
Nesi et al,[34] 2022	341	Any	Internalizing	0.12 (0.01 to 0.22)
Kelly et al,[41] 2019	1722	Any	Depression	0.15 (0.10 to 0.20)
Fardouly et al,[37] 2020	49	YouTube	Depression	0.19 (−0.10 to 0.45)
Model with RVE for clinical samples				0.08 (0.01 to 0.15)
Model with RVE for community samples				0.12 (0.09 to 0.15)



Forest plot for the individual and pooled effect sizes representing the association between time spent on social media and mental health symptoms. Effect sizes for clinical samples are shown both individually (ie, separate rows with author, year, and sample size) and as a pooled estimate (model with cluster-robust variance estimation [RVE]), while the effect size for community samples is only presented as a pooled estimate at the bottom. Individual Pearson r coefficients are depicted as filled squares with the size indicating the relative weight, based on sample size, of each effect size estimate for clinical studies in the meta-analysis. Increased time spent on social media was associated with decreased symptoms (to the left of zero) or increased symptoms (to the right of zero). The blue diamond and the dashed line represent the overall summary effect size across all clinical studies (r, 0.08; 95% CI, 0.01 to 0.15; P < .001), calculated using RVE to account for dependencies between effect sizes coming from the same study. The orange diamond and dashed line represent the overall summary effect size across all studies run on community samples (r, 0.12; 95% CI, 0.09 to 0.15; P < .001). The error bars and diamond width represent the 95% CIs for the effect sizes. The dotted reference line at r = 0 represents the point of reference for no correlation.

visual inspection of the funnel plot (eAppendix 11 in Supplement 1).

Sample type (clinical vs community) was not considered a significant moderator of the overall association between social media engagement and internalizing symptoms (β, 0.01; SE, 0.02; t, 0.72; 95% CI, −0.05 to 0.08; P = .52), and heterogeneity remained high (I^2, 94.70) (Table). Our additional moderation analyses, summarized in the Table, showed that neither mental health measure (anxiety vs internalizing: β, 0.03; SE, 0.04; t, 0.78; 95% CI, −0.11 to 0.17; P = .49; depression vs internalizing: β, 0.02; SE, 0.03; t, 0.74; 95% CI, −0.08 to 0.12; P = .53) nor COVID-19 (β, −0.06; SE, 0.05; t, −1.23; 95% CI, −0.15 to 0.04; P = .24) explained heterogeneity in the meta-correlation between social media engagement and mental health symptoms. There was also no moderation for the type of social media measures (Table), age, or sex (eAppendix 14 in Supplement 1).

Discussion

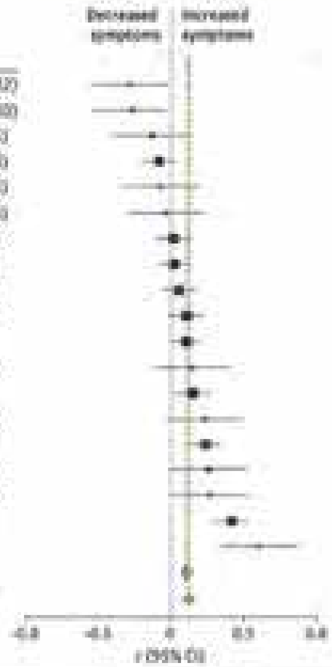
This systematic review and meta-analysis synthesized data from 16 years of research examining the association between social media use and internalizing mental health in more than 1 million adolescents. We found that 11% of studies examined clinical populations, while 88% recruited adolescents from the general population. There was a small, positive, and significant meta-correlation between social media use and internalizing mental health in clinical samples, regardless of whether time- or engagement-based social media metrics were studied. Notably, these meta-correlations did not substantially differ from those found in community samples.

Our first finding highlights a lack of research on clinical populations. Notably, adolescents affected by clinical-level anxiety and depression can face higher social withdrawal, sleep problems, low self-esteem, increased susceptibility to peer influence, and excessive rumination compared to adolescents from the general population.[45] These symptoms may alter their social media interaction and its impact on their mental health.[16,46,47] Hence, the lack of an evidence base in these high-risk populations, resulting in limited investigation of clinically relevant mechanisms, limits our capacity to draw accurate inferences about the relationship between social media use and mental health.

In contrast to the common assumption that clinical populations might show a stronger association between social media use and mental health declines than community samples,[17] we found no substantial differences. This result could be explained by the increasing occurrence of clinically significant symptoms in community samples[3,4] and the diminishing divide between these groups. Alternatively, adolescent clinical populations might adjust their social media use based on their mental health needs, leading to comparable usage patterns and correlations. Lastly, clinical groups could also be experienc-

Figure 4. Social Media Engagement and Internalizing Symptoms in Clinical Groups

Source	Sample, No.	Platform[a]	Mental health	r (95% CI)
Fardouly et al,[37] 2020	48	Activity	Anxiety	−0.30 (−0.54 to −0.02)
Fardouly et al,[37] 2020	48	Activity	Anxiety	−0.28 (−0.52 to −0.00)
Fardouly et al,[37] 2020	48	Activity	Anxiety	−0.14 (−0.41 to 0.15)
Reed et al,[33] 2021	243	Perception	Internalizing	−0.09 (−0.21 to 0.04)
Fardouly et al,[37] 2020	48	Activity	Depression	−0.08 (−0.35 to 0.21)
Fardouly et al,[37] 2020	48	Activity	Depression	−0.04 (−0.31 to 0.25)
Reed et al,[33] 2021	243	Perception	Internalizing	0.02 (−0.11 to 0.15)
Reed et al,[34] 2022	343	Perception	Internalizing	0.03 (−0.08 to 0.14)
Uçar et al,[39] 2018	224	Perception	Anxiety	0.06 (−0.07 to 0.19)
Reed et al,[33] 2021	243	Perception	Internalizing	0.12 (−0.01 to 0.24)
Reed et al,[34] 2022	343	Perception	Internalizing	0.12 (0.01 to 0.22)
Fardouly et al,[37] 2020	48	Activity	Depression	0.17 (−0.12 to 0.43)
Uçar et al,[39] 2018	224	Perception	Depression	0.17 (0.04 to 0.29)
Fardouly et al,[37] 2020	48	Activity	Anxiety	0.25 (−0.03 to 0.50)
Reed et al,[36] 2021	343	Perception	Internalizing	0.26 (0.16 to 0.36)
Fardouly et al,[37] 2020	48	Activity	Depression	0.28 (−0.00 to 0.52)
Fardouly et al,[37] 2020	48	Activity	Depression	0.28 (−0.00 to 0.52)
Reed et al,[33] 2021	243	Perception	Internalizing	0.43 (0.32 to 0.53)
Fardouly et al,[37] 2020	48	Activity	Anxiety	0.58 (0.36 to 0.74)
Model with RVE for clinical samples				0.12 (0.09 to 0.15)
Model with RVE for community samples				0.14 (0.10 to 0.18)



Forest plot for the individual and pooled effect sizes representing the association between social media engagement and mental health symptoms. Effect sizes for clinical samples are shown both individually (ie, separate rows with author, year, and sample size) and as a pooled estimate (model with cluster-robust variance estimation [RVE]), while the effect size for community samples is only presented as a pooled estimate at the bottom. Individual Pearson r coefficients are depicted as filled squares with the size indicating the relative weight, based on sample size, of each effect size estimate for clinical studies in the meta-analysis. Increased social media engagement was associated with decreased symptoms (to the left of zero) or increased symptoms (to the right of zero). The blue diamond and dashed line represent the overall summary effect size across all clinical studies (r, 0.12; 95% CI, 0.09 to 0.15; P < .001), calculated using RVE to account for dependencies between effect sizes coming from the same study. The error bars and diamond width represent the 95% CIs. The orange diamond and dashed line represent the overall summary effect size in community samples (r, 0.14; 95% CI, 0.10 to 0.18; P < .001). The dotted reference line at r = 0 represents the point of reference for no correlation. More information on the type of social media engagement measured in each study is reported in eAppendix 15 in Supplement 1.

[a] All reported measures of social media use are for any platform (no study measured activity or user perception in relation to a specific platform).

ing less variability in mental health symptomatology (eg, ceiling effects), lessening the observable correlations between social media use and mental health symptoms.

Limitations

We underscore some limitations of this work. First, inaccurate self-report measures of social media use[44] might have decreased our ability to locate differences between clinical and community samples even if they existed. Second, while we summarized studies with longitudinal effect sizes and selected control variables as part of our systematic review (eAppendix 11 in Supplement 1), our meta-analysis only included correlations. Hence, no causal inferences can be drawn from the pooled meta-correlation about whether increased social media use leads to higher symptoms or vice versa.

Further, we categorized social media engagement with 5 predefined categories, which are not exhaustive and could mask important nuances. For example, the role of social media content will depend on its nature, which could be positive, negative, or neutral. In addition, we focused on internalizing mental health only. Hence, conclusions cannot be generalized to other conditions. Limiting the inclusion of studies to those published in English may introduce language bias and exclude valuable research conducted in other languages.

Conclusions

The findings in this study demonstrated the moderate to high levels of heterogeneity common to this research area.[23] This variation could potentially be explained by individual differences in demographic characteristics among participants that we did not test, due to the lack of data or statistical power. However, when conducting exploratory moderation analyses for age and sex, we found that neither of those factors explained heterogeneity. We also found no evidence of publication bias.

Many worry about social media's role in increased clinical-level mental health symptoms among adolescents. However, current research falls short of adequately targeting the specific populations required to draw accurate inferences about this matter. Despite our initial findings of a similar association across clinical and community samples, there is still a real risk that we are incorrectly generalizing results from the gen-

Table. Moderation Analyses[a]

Moderator	Level	Studies, No.	Effect sizes, No.	Estimate (SE)	t Value	95% CI	P value
Time spent on social media and internalizing symptoms							
Sample type	Clinical (reference)	7	15	NA	NA	NA	NA
	Community	49	99	0.05 (0.03)	1.6	−0.02 to 0.12	.15
COVID-19	Before (reference)	44	100	NA	NA	NA	NA
	During	9	12	0.04 (0.04)	1.16	−0.04 to 0.13	.27
Mental health measure	Internalizing (reference)	4	5	NA	NA	NA	NA
	Depression	48	69	−0.07 (0.08)	−0.81	−0.30 to 0.17	.47
	Anxiety	19	40	−0.07 (0.08)	−0.81	−0.27 to 0.14	.45
Social media engagement and internalizing symptoms							
Sample type	Clinical (reference)	4	19	NA	NA	NA	NA
	Community	62	217	0.01 (0.02)	0.72	−0.05 to 0.80	.52
COVID-19	Before (reference)	53	196	NA	NA	NA	NA
	During	10	31	−0.06 (0.05)	−1.23	−0.15 to 0.04	.24
Mental health measure	Internalizing (reference)	3	8	NA	NA	NA	NA
	Depression	55	169	0.02 (0.03)	0.74	−0.08 to 0.12	.53
	Anxiety	24	58	0.03 (0.04)	0.78	−0.11 to 0.17	.49
Social media measure	Other (reference)	13	20	NA	NA	NA	NA
	Active vs passive	10	42	0.00 (0.04)	0.02	−0.09 to 0.09	.99
	Activity	22	69	0.05 (0.05)	0.90	−0.07 to 0.17	.38
	Content	2	25	−0.07 (0.04)	−1.74	−0.34 to 0.19	.27
	Perception	30	80	0.10 (0.05)	2.06	−0.00 to 0.20	.06

Abbreviation: NA, not applicable.

[a] Results of moderation analyses for the meta-correlation of internalizing symptoms with time spent on social media and engagement-based social media use measures.

eral population to young people with mental health conditions. The potential impact of this extends beyond research to clinical practice and policymaking. For clinicians, more research on clinical populations could enrich strategies for patient consultations and family education, allowing for the integration of social media management into treatment plans. For policymakers, it could shape policies for safer social media platforms and funding allocation toward mental health programs. In a world increasingly saturated by digital technology, we cannot afford to design prevention programs, interventions, and regulations without knowing that they work for everyone, especially those who are most vulnerable.

ARTICLE INFORMATION

Accepted for Publication: May 8, 2024.

Published Online: June 24, 2024. doi:10.1001/jamapediatrics.2024.2078

Author Contributions: Ms Fassi and Thomas had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.
Concept and design: Fassi, Thomas, Parry, Ford, Orben.
Acquisition, analysis, or interpretation of data: Fassi, Thomas, Parry, Leyland-Craggs.
Drafting of the manuscript: Fassi, Orben.
Critical review of the manuscript for important intellectual content: All authors.
Statistical analysis: Fassi, Parry.
Obtained funding: Orben.
Administrative, technical, or material support: Leyland-Craggs, Orben.
Supervision: Ford, Orben.

Conflict of Interest Disclosures: Dr Thomas reported grants from Wellcome Trust, Jacobs Foundation, UK Research and Innovation Economic and Social Research Council, and Gonville Studentship during the conduct of the study and from Wellcome Trust and Jacobs Foundation outside the submitted work. Dr Ford reported funding from Place2Be (research consultancy paid to research group) outside the submitted work. Dr Orben reported grants from UK Research and Innovation Medical Research Council, Jacobs Foundation, Emmanuel College, University of Cambridge, and UK Research and Innovation Future Leaders Fellowship during the conduct of the study as well as personal fees from Apple (honoraria for giving a talk at Apple University) outside the submitted work. No other disclosures were reported.

Funding/Support: The study received funding support through the Medical Research Council (RG86932, Ms Fassi; MC_UU_00030/13, Dr Orben), Wellcome Trust (WT104496/Z/15/Z, Ms Thomas), Stellenbosch University (Dr Parry), Jacobs Foundation (Ms Leyland-Craggs and Dr Orben), National Institute for Health and Care (Dr Ford), Cambridge Biomedical Research (Dr Ford), National Institute for Health and Care Applied Research Centre (Dr Ford), Place2Be (Dr Ford), Emmanuel College (Dr Orben), and UK Research and Innovation (MR/X034925/1, Dr Orben).

Role of the Funder/Sponsor: The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Data Sharing Statement: See Supplement 2.

Additional Contributions: We thank all the authors of the original studies that contributed to this systematic review and meta-analysis, and particularly those that responded to our requests regarding their research studies.

REFERENCES

1. Sawyer SM, Azzopardi PS, Wickremarathne D, Patton GC. The age of adolescence. *Lancet Child Adolesc Health*. 2018;2(3):223-228. doi:10.1016/S2352-4642(18)30022-1

2. NHS Digital. Rate of mental disorders among 17 to 19 year olds increased in 2022, new report shows. Accessed May 20, 2024. https://digital.nhs.uk/news/2022/rate-of-mental-disorders-among-17-to-19-year-olds-increased-in-2022-new-report-shows

3. NHS Digital. Mental health of children and young people in England, 2017. Accessed May 20, 2024. https://digital.nhs.uk/data-and-information/publications/statistical/mental-health-of-children-and-young-people-in-england/2017/2017

4. World Health Organization. Mental health. Published 2021. Accessed May 20, 2024. https://www.who.int/health-topics/mental-health

5. Ofcom. Ofcom Report 2022. https://www.ofcom.org.uk/__data/assets/pdf_file/0024/234609/childrens-media-use-and-attitudes-report-2022.pdf

6. Gunnell D, Kidger J, Elvidge H. Adolescent mental health in crisis. BMJ. 2018;361:k2608. doi:10.1136/bmj.k2608

7. Twenge JM, Joiner TE, Rogers ML, Martin GN. Corrigendum: "Increases in depressive symptoms, suicide-related outcomes, and suicide rates among US adolescents after 2010 and links to increased new media screen time." Clin Psychol Sci. 2019;7(2):397. doi:10.1177/2167702618824060

8. Orben A, Przybylski AK. The association between adolescent well-being and digital technology use. Nat Hum Behav. 2019;3(2):173-182. doi:10.1038/s41562-018-0506-1

9. Marciano L, Driver CC, Schulz PJ, Camerini AL. Dynamics of adolescents' smartphone use and well-being are positive but ephemeral. Sci Rep. 2022;12(1):1316. doi:10.1038/s41598-022-05291-y

10. Huang C. Time spent on social network sites and psychological well-being: a meta-analysis. Cyberpsychol Behav Soc Netw. 2017;20(6):346-354. doi:10.1089/cyber.2016.0758

11. Steinsbekk S, Nesi J, Wichstrøm L. Social media behaviors and symptoms of anxiety and depression. a four-wave cohort study from age 10-16 years. Comput Hum Behav. 2023;147:107859. doi:10.1016/j.chb.2023.107859

12. Odgers CL, Jensen MR. Annual research review: adolescent mental health in the digital age: facts, fears, and future directions. J Child Psychol Psychiatry. 2020;61(3):336-348. doi:10.1111/jcpp.13190

13. Valkenburg PM, Meier A, Beyens I. Social media use and its impact on adolescent mental health: an umbrella review of the evidence. Curr Opin Psychol. 2022;44:58-68. doi:10.1016/j.copsyc.2021.08.017

14. Ferguson CJ, Kaye LK, Branley-Bell D, et al. Like this meta-analysis: screen media and mental health. Prof Psychol Res Pract. 2022;53(2):205-214. doi:10.1037/pro0000426

15. Orben A, Przybylski AK, Blakemore SJ, Kievit RA. Windows of developmental sensitivity to social media. Nat Commun. 2022;13(1):1649. doi:10.1038/s41467-022-29296-3

16. Ghai S, Fassi L, Awadh F, Orben A. Lack of sample diversity in research on adolescent depression and social media use: a scoping review and meta-analysis. Clin Psychol Sci. 2023;11(5):759-772. doi:10.1177/21677026221114859

17. Liu M, Kamper-DeMarco KE, Zhang J, Xiao J, Dong D, Xue P. Time spent on social media and risk of depression in adolescents: a dose-response meta-analysis. Int J Environ Res Public Health. 2022;19(9):5164. doi:10.3390/ijerph19095164

18. Beyens I, Pouwels JL, van Driel II, Keijsers L, Valkenburg PM. The effect of social media on well-being differs from adolescent to adolescent. Sci Rep. 2020;10(1):10763. doi:10.1038/s41598-020-67727-7

19. Kieling C, Baker-Henningham H, Belfer M, et al. Child and adolescent mental health worldwide: evidence for action. Lancet. 2011;378(9801):1515-1525. doi:10.1016/S0140-6736(11)60827-1

20. Rao PA, Beidel DC, Turner SM, Ammerman RT, Crosby LE, Sallee FR. Social anxiety disorder in childhood and adolescence: descriptive psychopathology. Behav Res Ther. 2007;45(6):1181-1191. doi:10.1016/j.brat.2006.07.015

21. Teachman BA, Allen JP. Development of social anxiety: social interaction predictors of implicit and explicit fear of negative evaluation. J Abnorm Child Psychol. 2007;35(1):63-78. doi:10.1007/s10802-006-9084-1

22. Kostyrka-Allchorne K, Stoilova M, Bourgaize J, Rahali M, Livingstone S, Sonuga-Barke E. Review: Digital experiences and their impact on the lives of adolescents with pre-existing anxiety, depression, eating and nonsuicidal self-injury conditions—a systematic review. Child Adolesc Ment Health. 2023;28(1):22-32. doi:10.1111/camh.12619

23. Lilienfeld SO. Comorbidity between and within childhood externalizing and internalizing disorders: reflections and directions. J Abnorm Child Psychol. 2003;31(3):285-291. doi:10.1023/A:1023229529866

24. Protogerou C, Hagger MS. A checklist to assess the quality of survey studies in psychology. Methods Psychol. 2020;3:100031. doi:10.1016/j.metip.2020.100031

25. OSF. RMarkdown files for the systematic review and meta analysis. Accessed May 20, 2024. https://osf.io/gr3xh/?view_only=7b7434c46ad4468f9bea1cd1483f3e03

26. Wikimedia. International Telecommunications Union classification (2021). List of countries by regional classification. Accessed May 29, 2024. https://meta.wikimedia.org/wiki/List_of_countries_by_regional_classification

27. Fisher Z, Tipton E. robumeta: an R-package for robust variance estimation in meta-analysis. arXiv. Preprint posted online March 2015. https://arxiv.org/abs/1503.02220v1

28. Hedges LV, Tipton E, Johnson MC. Robust variance estimation in meta-regression with dependent effect size estimates. Res Synth Methods. 2010;1(1):39-65. doi:10.1002/jrsm.5

29. R Project. Robumeta package. Accessed May 20, 2024. https://cran.r-project.org/web/packages/robumeta/index.html

30. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ. 1997;315(7109):629-634. doi:10.1136/bmj.315.7109.629

31. Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. J Clin Epidemiol. 2008;61(10):991-996. doi:10.1016/j.jclinepi.2007.11.010

32. Viechtbauer W. Conducting meta-analyses in R with the metafor package. J Stat Softw. 2010;36(3):1-48. doi:10.18637/jss.v036.i03

33. Xiang M, Zhang Z, Kuwahara K. Impact of COVID-19 pandemic on children and adolescents' lifestyle behavior larger than expected. Prog Cardiovasc Dis. 2020;63(4):531-532. doi:10.1016/j.pcad.2020.04.013

34. Marques de Miranda D, da Silva Athanasio B, Sena Oliveira AC, Simoes-E-Silva AC. How is COVID-19 pandemic impacting mental health of children and adolescents? Int J Disaster Risk Reduct. 2020;51:101845. doi:10.1016/j.ijdrr.2020.101845

35. Salzano G, Passanisi S, Pira F, et al. Quarantine due to the COVID-19 pandemic from the perspective of adolescents: the crucial role of technology. Ital J Pediatr. 2021;47(1):40. doi:10.1186/s13052-021-00997-7

36. Nesi J, Burke TA, Extein J, et al. Social media use, sleep, and psychopathology in psychiatrically hospitalized adolescents. J Psychiatr Res. 2021;144:296-303. doi:10.1016/j.jpsychires.2021.10.034

37. Nesi J, Burke TA, Caltabiano A, Spirito A, Wolff JC. Digital media-related precursors to psychiatric hospitalization among youth. J Affect Disord. 2022;310:235-240. doi:10.1016/j.jad.2022.05.013

38. Fardouly J, Magson NR, Rapee RM, Johnco CJ, Oar EL. The use of social media by Australian preadolescents and its links with mental health. J Clin Psychol. 2020;76(7):1304-1326. doi:10.1002/jclp.22936

39. Hamilton JL, Biernesser C, Moreno MA, et al. Social media use and prospective suicidal thoughts and behaviors among adolescents at high risk for suicide. Suicide Life Threat Behav. 2021;51(6):1203-1212. doi:10.1111/sltb.12801

40. Uçar HN, Eray Ş, Kocael Ö, et al. Big data in adolescent psychiatry: do patients share their psychiatric symptoms on social networking sites? Psychiatr Danub. 2018;30(4):395-403. doi:10.24869/psyd.2018.395

41. Noel JK, Jacob S, Wensley IA, Rosenthal SR. Subjective smartphone screen time and co-morbid mental illness. J Technol Behav Sci. 2022;7(4):578-587. doi:10.1007/s41347-022-00276-0

42. Kelly Y, Zilanawala A, Booker C, Sacker A. Social media use and adolescent mental health: findings from the UK Millennium Cohort Study. EClinicalMedicine. 2018;6:59-68. doi:10.1016/j.eclinm.2018.12.005

43. Parry DA, Fisher JT, Mieczkowski H, Sewall CJR, Davidson BI. Social media and well-being: a methodological perspective. Curr Opin Psychol. 2022;45:101285. doi:10.1016/j.copsyc.2021.11.005

44. Meier A, Reinecke L. Computer-mediated communication, social media, and mental health: a conceptual and empirical meta-review. Commun Res. 2020;48(8). doi:10.1177/0093650220958224

45. Huberty TJ. Emotional and behavioral problems, students with. In: Spielberger CD, ed. Encyclopedia of Applied Psychology. Elsevier; 2004:723-730. doi:10.1016/B0-12-657410-3/00791-1

46. Radovic A, Gmelin T, Stein BD, Miller E. Depressed adolescents' positive and negative use of social media. J Adolesc. 2017;55:5-15. doi:10.1016/j.adolescence.2016.12.002

47. Weinstein E, Kleiman EM, Franz PJ, et al. Positive and negative uses of social media among adolescents hospitalized for suicidal behavior. J Adolesc. 2021;87:63-73. doi:10.1016/j.adolescence.2020.12.003

48. Scharkow M. The accuracy of self-reported internet use—a validation study using client log data. Commun Methods Meas. 2016;10(1):13-27. doi:10.1080/19312458.2015.1118446

Journal of Adolescent Health 72 (2023) S52–S60

JOURNAL OF
ADOLESCENT
HEALTH
www.jahonline.org

Original article

Detecting Depression and Anxiety Among Adolescents in South Africa: Validity of the isiXhosa Patient Health Questionnaire-9 and Generalized Anxiety Disorder-7

Marguerite Marlow, M.A.[a,*], Sarah Skeen, Ph.D.[a,b], Caitlin M. Grieve, BA Hons.[a], Liliana Carvajal-Velez, M.Sc.[c,d], Jill W. Åhs, M.Med.Sci.[e,f], Brandon A. Kohrt, M.D., Ph.D.[g], Jennifer Requejo, Ph.D.[c], Jackie Stewart, Ph.D.[h], Junita Henry, M.A.[a], Daniel Goldstone, M.A.[a], Tashmira Kara, BA Hons.[a], and Mark Tomlinson, Ph.D.[a,i]

[a] Institute for Life Course Health Research, Department of Global Health, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa
[b] Amsterdam Institute for Social Science Research, University of Amsterdam, Amsterdam, Netherlands
[c] Division of Data, Analytics, Planning and Monitoring, Data and Analytics Section, UNICEF, New York, New York
[d] Department of Global Public Health, Karolinska Institutet, Stockholm, Sweden
[e] Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden
[f] Department of Health Sciences, Swedish Red Cross University College, Stockholm, Sweden
[g] Department of Psychiatry, George Washington University, Washington, District of Columbia
[h] Division of Global Surgery, University of Cape Town, Cape Town, South Africa
[i] School of Nursing and Midwifery, Queens University, Belfast, United Kingdom

Article history: Received April 11, 2022; Accepted September 27, 2022
Keywords: Adolescents; Mental health; Depression; Anxiety; Screening; PHQ-9;
middle-income countries

Sx's
vs dx

A B S T R A C T

Purpose: Screening tools such as the Patient Health Questionnaire-9 (PHQ-9) an
Anxiety Disorder-7 (GAD-7) could potentially be used in resource-limited setting
lescents who need mental health support. We examined the criterion validity of
sions of the PHQ-9 and GAD-7 in detecting depression and anxiety among adolesce
in South Africa.
Methods: Adolescents were recruited from the general population and from
organizations working with adolescents in need of mental health support. The P
were culturally adapted and translated into isiXhosa and administered to 302 adolescents (56.8% female). The Kiddie Schedule for Affective Disorders and Schizophrenia was administered by trained clinicians as the gold standard diagnostic measure for depression and anxiety.
Results: For the PHQ-9, the area under the curve was 0.88 for the full sample of adolescents (10–19 years old). A score of ≥10 had 91% sensitivity and 76% specificity for detecting adolescents with depression. For the GAD-7, the area under the curve was 0.78, and cutoff scores with an optimal sensitivity-specificity balance were low (≥6). A score of ≥6 had 67% sensitivity and 75% specificity for detecting adolescents with anxiety.

Generalized Anxiety Disorder-7 for a range of cutoff scores, for use with adolescents. These findings make a meaningful contribution to establishing tools to measure adolescent mental health at a population level in South Africa and other

Conflicts of interest: The authors have no conflicts of interest or competing interests relevant to this work. The funder (Bill & Melinda Gates Foundation) had no role to play in the study design, data collection, analysis, or manuscript preparation.
Disclaimer: The article was published as part of supplement that was supported by the Bill & Melinda Gates Foundation [INV-001395] and UNICEF.

* Address correspondence to: Marguerite Marlow, Institute for Life Course Health Research, Department of Global Health, Faculty of Medicine and Health Sciences, Stellenbosch University, Francie van Zijl Drive, Tygerberg, Cape Town, Western Cape 7505, South Africa.
E-mail address: mmarlow@sun.ac.za (M. Marlow).

https://doi.org/10.1016/j.jadohealth.2022.09.013

Journal of Adolescent Health 72 (2023) S52–S60

ELSEVIER

JOURNAL OF
ADOLESCENT
HEALTH

www.jahonline.org

Original article

Detecting Depression and Anxiety Among Adolescents in South Africa: Validity of the isiXhosa Patient Health Questionnaire-9 and Generalized Anxiety Disorder-7

Marguerite Marlow, M.A. [a,*], Sarah Skeen, Ph.D. [a,b], Caitlin M. Grieve, BA Hons. [a], Liliana Carvajal-Velez, M.Sc. [c,d], Jill W. Åhs, M.Med.Sci. [e,f], Brandon A. Kohrt, M.D., Ph.D. [g], Jennifer Requejo, Ph.D. [c], Jackie Stewart, Ph.D. [h], Junita Henry, M.A. [a], Daniel Goldstone, M.A. [a], Tashmira Kara, BA Hons. [a], and Mark Tomlinson, Ph.D. [a,i]

[a] *Institute for Life Course Health Research, Department of Global Health, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa*
[b] *Amsterdam Institute for Social Science Research, University of Amsterdam, Amsterdam, Netherlands*
[c] *Division of Data, Analytics, Planning and Monitoring, Data and Analytics Section, UNICEF, New York, New York*
[d] *Department of Global Public Health, Karolinska Institutet, Stockholm, Sweden*
[e] *Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden*
[f] *Department of Health Sciences, Swedish Red Cross University College, Stockholm, Sweden*
[g] *Department of Psychiatry, George Washington University, Washington, District of Columbia*
[h] *Division of Global Surgery, University of Cape Town, Cape Town, South Africa*
[i] *School of Nursing and Midwifery, Queens University, Belfast, United Kingdom*

Article history: Received April 11, 2022; Accepted September 27, 2022
Keywords: Adolescents; Mental health; Depression; Anxiety; Screening; PHQ-9; GAD-7; Validity; Diagnostic assessment; Low- and middle-income countries

A B S T R A C T

Purpose: Screening tools such as the Patient Health Questionnaire-9 (PHQ-9) and the Generalized Anxiety Disorder-7 (GAD-7) could potentially be used in resource-limited settings to identify adolescents who need mental health support. We examined the criterion validity of the isiXhosa versions of the PHQ-9 and GAD-7 in detecting depression and anxiety among adolescents (10–19 years) in South Africa.
Methods: Adolescents were recruited from the general population and from nongovernmental organizations working with adolescents in need of mental health support. The PHQ-9 and GAD-7 were culturally adapted and translated into isiXhosa and administered to 302 adolescents (56.9% female). The Kiddie Schedule for Affective Disorders and Schizophrenia was administered by trained clinicians as the gold standard diagnostic measure for depression and anxiety.
Results: For the PHQ-9, the area under the curve was 0.88 for the full sample of adolescents (10–19 years old). A score of ≥10 had 91% sensitivity and 76% specificity for detecting adolescents with depression. For the GAD-7, the area under the curve was 0.78, and cutoff scores with an optimal sensitivity-specificity balance were low (≥6). A score of ≥6 had 67% sensitivity and 75% specificity for detecting adolescents with anxiety.

IMPLICATIONS AND CONTRIBUTIONS

This study determined the psychometric properties of the culturally adapted isiXhosa versions of the Patient Health Questionnaire-9 and Generalized Anxiety Disorder-7 for a range of cutoff scores, for use with adolescents. These findings make a meaningful contribution to establishing tools to measure adolescent mental health at a population level in South Africa and other

Conflicts of interest: The authors have no conflicts of interest or competing interests relevant to this work. The funder (Bill & Melinda Gates Foundation) had no role to play in the study design, data collection, analysis, or manuscript preparation.
Disclaimer: The article was published as part of supplement that was supported by the Bill & Melinda Gates Foundation (INV-001395) and UNICEF.

* Address correspondence to: Marguerite Marlow, Institute for Life Course Health Research, Department of Global Health, Faculty of Medicine and Health Sciences, Stellenbosch University, Francie van Zijl Drive, Tygerberg, Cape Town, Western Cape 7505, South Africa.
E-mail address: mmarlow@sun.ac.za (M. Marlow).

https://doi.org/10.1016/j.jadohealth.2022.09.013

Discussion: The culturally adapted isiXhosa version of the PHQ-9 can be used as a valid measure for depression in adolescents. Further research on the GAD-7 for use with adolescents is recommended.

low- and middle-income countries.

Adolescence is a particularly vulnerable period for mental health, with almost half of all mental disorders developing before the age of 18 years [1]. Although major commitments have been made to improve mental health research globally [2], we still know little about the mental health of the majority of the world's children and adolescents [3]. Better prevalence data from low- and middle-income countries (LMICs) are urgently needed, particularly from sub-Saharan Africa where data coverage is essentially nonexistent [3]. Research on adolescent mental health in LMICs is hampered by the lack of validated measurement tools that can be used at a population level. A recent systematic review reported high rates of depression (26.9%) and anxiety (29.8%) among the general population of adolescents in sub-Saharan Africa [4]. However, most studies used screening tools, many of which have not been validated for use with adolescents in these settings. Cultural adaptation of tools and appropriate validation efforts in LMICs are needed to ensure that reported prevalence rates do not underestimate or overestimate the burden of the problem [5]. Accurate data on prevalence are important to ensure that resources for mental health services are appropriately allocated [6].

This study was conducted as part of the United Nations Children's Fund's Measurement of Mental Health Among Adolescents at the Population Level (MMAP) initiative that aims to develop and validate measurement tools that support large-scale collection of robust data on adolescent mental health [7]. To allow for effective integration into national data-collection efforts, measures need to be brief and easy to administer, using suitable language and phrasing for the population and setting. In addition, the tools should lend themselves to cross-cultural adaptation and for multicountry comparisons. Two of the measures being validated for adolescents as part of the MMAP initiative are the widely used Patient Health Questionnaire-9 (PHQ-9), a screening tool for depression, and the Generalized Anxiety Disorder-7 (GAD-7), a screening tool for anxiety. Both tools have been used with adolescents in a handful of sub-Saharan African countries [8–12], but they have not been validated for use among adolescents in South Africa.

Our study examined the criterion validity of the culturally adapted isiXhosa versions of the PHQ-9 and GAD-7 in detecting depression and anxiety among adolescents in Khayelitsha, South Africa.

Methods

Setting

The study was conducted in Khayelitsha, a large periurban neighborhood outside of Cape Town, South Africa. Khayelitsha, meaning "new home" in isiXhosa, was originally established for Black people under the Apartheid government. The area remains affected by inadequate service delivery and high rates of socioeconomic deprivation [13], with many households affected by domestic violence, assault, rape, and murder [14]. As such, adolescents are commonly exposed to high-stress living environments, placing them at greater risk of poor mental health.

Participants

Participants were isiXhosa-speaking adolescents between the ages of 10 and 19 years, living in Khayelitsha. To effectively assess the measures' psychometric stability [15], we aimed for a sample of 300 adolescents across the younger (10–14 years) and older (15–19 years) age range, with an even split between male and female adolescents. We aimed to obtain an enriched sample or a specific proportion of participants likely to have anxiety or depression at a 2:1 ratio to participants without these conditions. We conducted rolling recruitment to ensure that an adequate number of adolescents likely to have symptoms of depression or anxiety were included, using three methods of recruitment. First, we used school-based recruitment where the research team presented the project to one class (usually with 40–50 students) in each grade (grades 4–11). Information sheets were distributed to take home, and interested families were asked to return the completed form to the school. Second, we recruited older adolescents directly from existing community networks, such as street committees. Third, to increase the number of adolescents in the sample likely to have depression or anxiety, we recruited directly from local nongovernmental organizations that work with adolescents in need of mental health support.

Measures

Depression and anxiety screening. We used the isiXhosa versions of the PHQ-9 [16] and GAD-7 [17], which were adapted for use (Box 1). Both measures ask participants to rate how often over the last two weeks they were bothered by specific symptoms, with scores ranging from 0 (not at all) to 3 (nearly every day). For the PHQ-9, we used a slight variation to item 7 ("trouble concentrating on things like **school work**, reading, or watching television" as opposed to "...**reading the newspaper** or watching television"), similar to the PHQ-A (a modified version for adolescents).

Broader mental health. The PHQ-9 and GAD-7 were included as part of a larger MMAP questionnaire, consisting of United Nations Children's Fund-developed measures to assess suicidal ideation and behavior, functional limitations, mental health care, and connectedness [18].

Diagnostic assessment. We used the Kiddie Schedule for Affective Disorders and Schizophrenia (K-SADS) as the gold standard diagnostic measure for depression and anxiety.

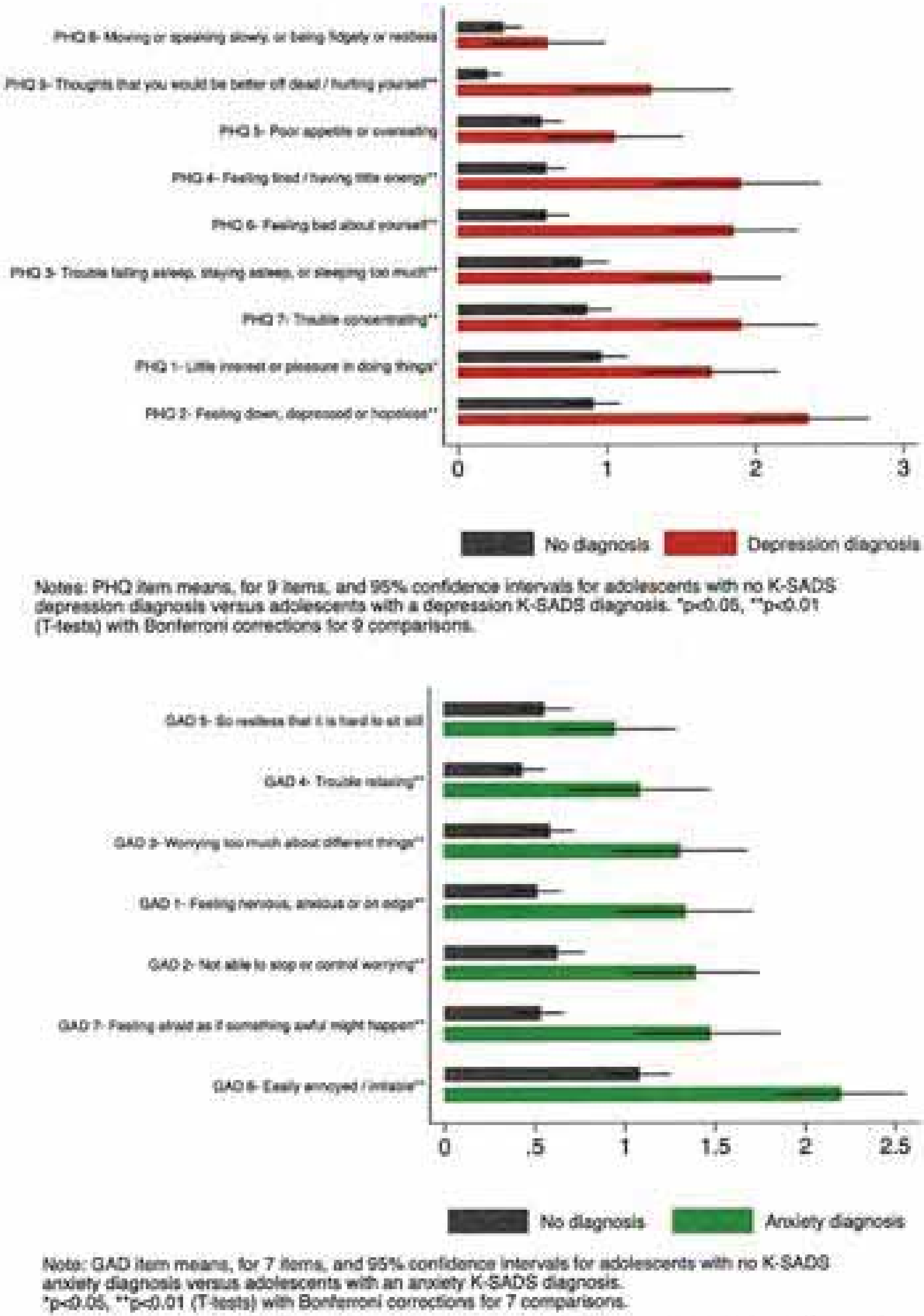


Figure 1. Endorsement of PHQ-9 items by the K-SADS depression diagnosis and endorsement of GAD-7 items by the K-SADS anxiety diagnosis (means and 95% CI).

Box 1. Transcultural translation and adaptation

We used a systematic transcultural translation and adaptation process [7] to produce an isiXhosa version of the Measurement of Mental Health Among Adolescents at the Population Level questionnaire. Original items were translated from English to isiXhosa by bilingual local mental health experts via group consensus, with a focus on cultural nuances and age-appropriate language. Translated items were presented to adolescents and their caregivers via focus group discussions (n = 6) and cognitive interviews (n = 18). Participant feedback was incorporated into an updated version of the questionnaire, which we presented to a group of local researchers for final revisions. Lastly, we conducted a blind back translation to ensure that translated items retained the original meaning. The adaptation process resulted in the following key changes:

(1) The response category of "several days" (a score of one on the Patient Health Questionnaire-9 and Generalized Anxiety Disorder-7) was changed in the translated version to "a few days";
(2) Items were reframed from statements ("*I feel...*") to questions ("*how often do you feel...?*");
(3) Wording of certain items was adapted to increase clarity for adolescents;
(4) Visual aids were produced to assist with respondents' understanding of time frames and response categories (Figure A1).

The adapted versions of Patient Health Questionnaire-9 and Generalized Anxiety Disorder-7 in English and isiXhosa are available in the Supplemental Annex.

The K-SADS is a semistructured interview for children aged 6–18 years, designed to assess ongoing psychopathology, including mood and anxiety disorders [19]. Criteria-based algorithms determine the presence of current disorders, in line with the Diagnostic and Statistical Manual of Mental Disorders [20]. We used the depression module and a selection of anxiety modules (generalized, social, and separation) that correspond with anxiety disorders that commonly impact this age group.

The K-SADS was translated into isiXhosa by mental health professionals fluent in both languages through a process of group consensus. We piloted the translated version with adolescents, under the supervision of a clinical psychologist, to refine the translations and to establish inter-rater reliability.

Procedures

The study protocol was approved by the Health Research Ethics Committee at Stellenbosch University (N19/08/104) and the Western Cape Department of Education (20200206-4132). The MMAP questionnaire was administered as an interview by local trained data collectors, fluent in English and isiXhosa, experienced in working with adolescents in Khayelitsha. The training focused on informed assent and consent procedures, questionnaire administration, data management, and referral and emergency protocols. The K-SADS was administered as the diagnostic assessment by trained and supervised social workers, fluent in isiXhosa and experienced in working with children and adolescents. Similar to many LMICs, South Africa is characterized by severe human resource shortages for mental health [21]. Specialists such as psychiatrists and psychologists—particularly those who speak local languages such as isiXhosa—are in exceptionally short supply. As a result, no isiXhosa-speaking psychologist, psychiatric nurse, or clinical social worker was available to conduct the diagnostic assessments for the study. Two social workers administered the K-SADS, under the supervision of a registered clinical psychologist (D.G.). They received extensive training over a three-week period, including training on the presentation and identification of depressive and anxiety disorders, interviewing techniques, K-SADS administration, and referrals.

A three-day pilot of assessment procedures took place to identify administration or logistical issues before starting the study. Six adolescents (three in the younger age group and three in the older age group) participated in the pilot.

Data collection took place at a community research center in Khayelitsha between May and November 2021. Following informed assent or consent, adolescents were interviewed in isiXhosa, using the MMAP questionnaire. Trained data collectors captured participant responses using a preprogrammed questionnaire on tablet devices. Alerts were programmed to flag the need to activate a referral protocol if participants were considered to be at risk. Interviews were audio-recorded for quality control purposes. After a refreshment break, adolescents proceeded to see a social worker for the K-SADS interview, with scores captured on tablet devices. Following the interview, the social worker and data collector met to discuss the need for referral, and the social worker facilitated referrals to relevant services. The entire assessment visit lasted 1.5–2 hours. Any K-SADS interviews that could not be completed on the same day as the MMAP interview were scheduled within a 48-hour window. As a token of appreciation, all participants received an R160 (~11 USD) supermarket voucher.

Online data submissions were reviewed daily to track progress and referrals. A random selection of audio-recordings from each data collector were reviewed weekly, with constructive feedback provided during team meetings. K-SADS assessments were video-recorded to enable a detailed review of cases and ongoing training during weekly supervision sessions.

Statistical analysis

Descriptive statistics were completed for the full sample and stratified by younger and older age groups. Based on the K-SADS results, the following diagnostic categories were used for the analysis: depression diagnosis (evidence of a major depressive disorder), anxiety diagnosis (evidence of generalized anxiety, separation anxiety, and/or social anxiety disorder), any diagnosis (evidence of depression and/or anxiety), and no diagnosis.

To assess the criterion validity of the PHQ-9 and GAD-7 against a clinician's diagnosis, we used total scores to construct a receiver operating characteristic curve and calculated the area under the curve (AUC) for each test, using the K-SADS as the gold standard. For AUC calculations, we report exact binomial 95% confidence intervals (95% CI). To make determinations related to the tests' validity, we calculated psychometric properties (Table A1) for a range of cutoff values. We completed all analyses using Stata, version 17 (Statcorp LP, College Station, Texas).

Table 1
Participant demographic information

	10–14 Years old (n = 130)	15–19 Years old (n = 172)	Total sample (n = 302)
Age			
Mean (SD)	12.09 (1.43)	17.02 (1.52)	14.90 (2.86)
Gender			
Male	50 (38.5%)	78 (45.3%)	128 (42.4%)
Female	80 (61.5%)	92 (53.5%)	172 (56.9%)
Other	0	2 (1.2%)	2 (0.7%)
Recruitment source			
School/Community	66 (50.8%)	67 (39%)	133 (44%)
NGO	64 (49.2%)	105 (61%)	169 (56%)
Home language			
isiXhosa	128 (98.5%)	167 (97.1%)	295 (97.7%)
English/other	2 (1.5%)	5 (2.9%)	7 (2.3%)
School language			
isiXhosa	93 (71.5%)	111 (64.5%)	204 (67.5%)
English/other	37 (28.5%)	61 (35.5%)	98 (32.5%)
Education (highest completed grade)[a]			
Grade 1–4	48 (37.2%)	0	48 (16%)
Grade 5–7	78 (60.5%)	18 (10.6%)	96 (32.1%)
Grade 8–10	3 (2.3%)	103 (60.6%)	106 (35.5%)
Grade 11–12	0	49 (28.8%)	49 (16.4%)
PHQ-9 (depression symptom screen)			
Mean (SD)	7.05 (5.28)	7.31 (5.19)	7.20 (5.22)
Moderate to severe symptoms (score of 10+)	39 (30%)	58 (33.7%)	97 (32.1%)
GAD-7 (anxiety symptom screen)			
Mean (SD)	4.28 (3.85)	5.63 (4.74)	5.05 (4.42)
Moderate to severe symptoms (score of 10+)	15 (11.5%)	39 (22.7%)	54 (17.8%)
Suicide attempt (ever)[b]	7 (5.4%)	30 (17.5%)	37 (12.3%)
Kiddie Schedule of Affective Disorders and Schizophrenia (K-SADS)			
No diagnosis	118 (90.7%)	127 (73.8%)	245 (81.1%)
Depression diagnosis	3 (2.3%)	20 (11.6%)	23 (7.6%)
Anxiety diagnosis	9 (6.9%)	36 (20.9%)	45 (14.9%)
Any diagnosis	12 (9.2%)	45 (26.1%)	57 (18.9%)

NGO = nongovernmental organization; SD = standard deviation.
[a] Missing: n = 3.
[b] Missing: n = 2.

Results

The sample consisted of 302 adolescents (56.9% female), with 43% younger and 57% older adolescents (Table 1). Over half of the sample (56%) were recruited from organizations working with adolescents in need of mental health support, and 44% were recruited via school or community avenues.

Based on severity thresholds of the original PHQ-9 [16], 32.1% of adolescents (65% female) were categorized with moderate to severe depression symptoms (a total score of 10 or higher). Using the GAD-7 [17], 17.8% (57% female) of adolescents were categorized with moderate to severe anxiety symptoms (a total score of 10 or higher). In addition, 12.3% of adolescents—predominantly older adolescents—reported that they had previously attempted suicide.

Using the K-SADS as the diagnostic tool, 7.6% of adolescents were diagnosed with depression, and 14.9% were diagnosed with anxiety. Eleven adolescents (3.6%) were diagnosed with both anxiety and depression.

During the assessment visit, six participants were identified as actively suicidal (had a plan and intended to carry out the plan) and were immediately referred to the nearest hospital for assistance.

Psychometric analysis

Results are presented for the full sample and for older adolescents (15–19 years old). Due to the small number of younger adolescents with a diagnosis according to the K-SADS (n = 3 for depression; n = 9 for anxiety), we were not able to make generalizable statements about the PHQ-9's or GAD-7's performance with this age group.

Using the PHQ-9 to discriminate between adolescents with and without a depression diagnosis on the K-SADS, the AUC (Figure A2) was 0.88 for the full sample (95% CI 0.81–0.95) and 0.88 for older adolescents (95% CI 0.80–0.96).

Using the GAD-7 to discriminate between adolescents with and without an anxiety diagnosis on the K-SADS, the AUC was 0.78 for the full sample (95% CI 0.71–0.85) and 0.79 for older adolescents (95% CI 0.71–0.87) (Figure A3).

Table 2 shows the test characteristics of the PHQ-9 and GAD-7 using the K-SADS as the gold standard. The PHQ-9 performed best for discriminating depression among the full sample, with the highest diagnostic odds ratio (OR) of 33.89 at a cutoff score of 10 or greater, with a sensitivity of 0.91, specificity of 0.76, and an overall accuracy of 0.77. For older adolescents, the highest diagnostic OR of 31.24 resulted from a cutoff score of 11 or greater. At this cutoff, the PHQ-9 had a sensitivity of 0.90 and a specificity of 0.78 for detecting older adolescents with major depression on the K-SADS, with an overall accuracy of 0.79.

Compared to the PHQ-9, validity was slightly weaker for the GAD-7: The optimal cutoff value for maximizing sensitivity without loss of specificity was a score of six or greater for older adolescents and for the full sample. For the full sample, a cutoff score of six had a sensitivity of 0.67 and a specificity of 0.75, with a diagnostic OR of 5.91 and an overall accuracy of 0.74. For older adolescents, a cutoff score of six had higher sensitivity (0.72) but lower specificity (0.68), with a slightly lower diagnostic OR of 5.62 and overall accuracy of 0.69.

Item analysis

Item means for the PHQ-9 and GAD-7 were calculated separately for adolescents who were diagnosed as having an anxiety or depressive disorder versus adolescents who did not receive any diagnosis. For the anxiety analysis, we excluded adolescents with depression who did not have anxiety. For the depression analysis, we excluded adolescents who had anxiety without comorbid depression. Table 3 presents the discriminant ability of items on the PHQ-9 and GAD-7 for adolescents with and without a diagnosis according to the K-SADS results. For the PHQ-9, the majority of items performed well by showing significant differences between respondents with and those without a depression diagnosis, with the exception of item 5 ("*poor appetite or overeating*") and item 8 ("*moving or speaking slowly/being fidgety or restless*"). For the GAD-7, item 5 ("*so restless that it is hard to sit still*") did not discriminate between diagnosed and undiagnosed groups.

Figure 1 provides PHQ-9 item means by K-SADS diagnostic status, comparing adolescents with no diagnosis to those with a depression diagnosis. The three most frequently endorsed items on the PHQ-9 among undiagnosed adolescents were item 1 (*Little interest or pleasure in doing things*), item 2 (*Feeling down, depressed, or hopeless*), and item 7 (*Trouble concentrating*). All three items performed well to distinguish between depressed and nondepressed adolescents.

Table 2
Validation psychometrics of the PHQ-9 and the GAD-7 from comparison with the Kiddie Schedule of Affective Disorders and Schizophrenia (K-SADS)

	Cutoff score	Sensitivity	Specificity	PPV	NPV	PLR	NLR	Diagnostic OR	Youden's Index (J)	TP (%)	TN (%)	FP (%)	FN (%)	Accurately classified (%)
Total PHQ-9 score														
Full sample	≥7	0.91	0.62	0.17	0.99	2.43	0.14	17.40	0.54	0.07	0.58	0.35	0.01	0.65
(10–19 years old)	≥8	0.91	0.67	0.18	0.99	2.74	0.13	21.00	0.58	0.07	0.62	0.31	0.01	0.69
	≥9	0.91	0.73	0.22	0.99	3.35	0.12	28.05	0.64	0.07	0.67	0.25	0.01	0.74
	≥10	0.91	0.76	0.24	0.99	3.86	0.11	33.89	0.68	0.07	0.71	0.22	0.01	0.77
	≥11	0.74	0.82	0.26	0.97	4.21	0.32	13.30	0.56	0.06	0.76	0.16	0.02	0.82
	≥12	0.74	0.88	0.34	0.98	6.25	0.30	21.12	0.62	0.06	0.81	0.11	0.02	0.87
	≥13	0.57	0.92	0.37	0.96	7.17	0.47	15.19	0.49	0.04	0.85	0.07	0.03	0.89
Older adolescents	≥7	0.90	0.55	0.21	0.98	1.98	0.18	10.83	0.45	0.10	0.48	0.40	0.01	0.59
(15–19 years old)	≥8	0.90	0.64	0.25	0.98	2.49	0.16	15.87	0.54	0.10	0.56	0.32	0.01	0.67
	≥9	0.90	0.67	0.26	0.98	2.74	0.15	18.36	0.57	0.10	0.59	0.29	0.01	0.70
	≥10	0.90	0.74	0.31	0.98	3.42	0.14	25.20	0.64	0.10	0.65	0.23	0.01	0.76
	≥11	0.90	0.78	0.35	0.98	4.02	0.13	31.24	0.68	0.10	0.69	0.20	0.01	0.79
	≥12	0.70	0.85	0.38	0.96	4.63	0.35	13.09	0.55	0.08	0.75	0.13	0.03	0.83
	≥13	0.70	0.89	0.47	0.96	6.65	0.34	19.83	0.59	0.08	0.79	0.09	0.03	0.87
Total GAD-7 score														
Full sample	≥4	0.82	0.65	0.29	0.95	2.32	0.28	8.44	0.47	0.12	0.55	0.30	0.03	0.67
(10–19 years old)	≥5	0.76	0.70	0.31	0.94	2.52	0.35	7.23	0.46	0.11	0.60	0.25	0.04	0.71
	≥6	0.67	0.75	0.32	0.93	2.64	0.45	5.91	0.41	0.10	0.64	0.22	0.05	0.74
	≥7	0.62	0.78	0.33	0.92	2.86	0.48	5.95	0.40	0.09	0.67	0.19	0.06	0.76
	≥8	0.58	0.84	0.38	0.92	3.54	0.50	7.01	0.41	0.09	0.71	0.14	0.06	0.80
	≥9	0.49	0.87	0.40	0.91	3.81	0.59	6.49	0.36	0.07	0.74	0.11	0.08	0.81
	≥10	0.40	0.91	0.44	0.90	4.47	0.66	6.78	0.31	0.06	0.77	0.08	0.09	0.83
Older adolescents	≥4	0.86	0.49	0.31	0.93	1.67	0.29	5.85	0.35	0.18	0.38	0.41	0.03	0.56
(15–19 years old)	≥5	0.81	0.63	0.36	0.92	2.15	0.31	6.90	0.43	0.17	0.49	0.30	0.04	0.66
	≥6	0.72	0.68	0.38	0.90	2.28	0.41	5.62	0.41	0.15	0.54	0.25	0.06	0.69
	≥7	0.67	0.74	0.40	0.89	2.52	0.45	5.56	0.40	0.14	0.58	0.21	0.07	0.72
	≥8	0.64	0.77	0.43	0.89	2.80	0.47	5.99	0.41	0.13	0.61	0.18	0.08	0.74
	≥9	0.64	0.82	0.48	0.90	3.48	0.44	7.86	0.46	0.13	0.65	0.15	0.08	0.78
	≥10	0.53	0.85	0.49	0.87	3.59	0.55	6.48	0.38	0.11	0.67	0.12	0.10	0.78

FN = false negative; FP = false positive; GAD-7 = Generalized Anxiety Disorder-7; NLR = negative likelihood ratio; NPV = negative predictive value; OR = odds ratio; PHQ-9 = Patient Health Questionnaire-9; PLR = positive likelihood ratio; PPV = positive predictive value; TN = true negative; TP = true positive.

Figure 1 also provides GAD-7 item means by K-SADS diagnostic status, comparing undiagnosed adolescents to those with an anxiety diagnosis. The three most frequently endorsed items on the GAD-7 among undiagnosed adolescents were item 6 (*Easily annoyed/irritable*), item 2 (*Not being able to stop or control worrying*), and item 3 (*Worrying too much about different things*). Similar to the PHQ-9, all three items successfully discriminated between diagnosed and undiagnosed adolescents.

Adjustments for population prevalence

Using cutoff values that balance sensitivity and specificity, Figure 2 shows what would be reflected in the outcomes of reported prevalence rates on the PHQ-9 and GAD-7 based on estimates of true prevalence rates. When adjusting for false positives and false negatives, the PHQ-9 and GAD-7 identified prevalence rates for depression and anxiety can be adjusted to approximate what the true prevalence may be in the population. For example, if a prevalence of 36% is identified for depression among 15- to 19-year-old adolescents, the estimated true prevalence is likely closer to 20%. The degree of adjustment differs based on prevalence because of the contribution of false positive versus false negatives to the estimates made. The positive predictive value and negative predictive value for individual adolescents also vary by prevalence rates (also included in Figure 2). Policy-makers can use algorithms or figures such as this to make adjusted prevalence estimates when allocating resources and designing programs.

Discussion

Our study is the first to evaluate the psychometric properties of the PHQ-9 and GAD-7 against a diagnostic interview for use with adolescents in South Africa. This validation exercise of the culturally adapted isiXhosa versions of the PHQ-9 and GAD-7 makes a meaningful contribution to establishing tools to measure adolescent mental health at a population level in South Africa and potentially other LMICs.

Our sample included adolescents from the general population as well as those attending nongovernmental organizations that provide adolescents with mental health support. We determined the psychometric properties of the PHQ-9 and GAD-7 for a range of cutoff scores. Cutoff scores should be selected based on the intended use of the tool for different applications. For example, the cutoff scores appropriate for screening adolescents to include in interventions may differ from cutoff scores used for determining population prevalence.

Measures with high sensitivity should be prioritized when risk factors to health and safety are serious [5]. For adolescent populations in LMICs, using tools with the highest possible sensitivity is crucial given that self-harm is among the top five causes of death and that depressive and anxiety disorders are among the leading causes of disability [22]. At the same time, high specificity is also important in population-based research, to reduce the likelihood of overburdening resource-deprived health systems with high numbers of false positives. Achieving high specificity without compromising sensitivity is therefore a key priority for measuring adolescent mental health at a

Table 3
Discriminant ability of PHQ-9 and GAD-7 items for adolescents with and without diagnosis on the Kiddie Schedule of Affective Disorders and Schizophrenia (K-SADS) for the full sample and subsample of 15–19 years old

PHQ-9		15–19 Years old		Full sample	
		Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Item #	Description	No diagnosis (n = 127)	Depression (n = 20)	No diagnosis (n = 245)	Depression (n = 23)
PHQ 1	Little interest or pleasure in doing things	0.96 (0.99)	1.70 (1.03)*	1.11 (1.09)	1.78 (1.04)*
PHQ 2	Feeling down, depressed, or hopeless	0.91 (1.03)	2.35 (0.93)*	0.83 (0.99)	2.39 (0.89)*
PHQ 3	Trouble falling asleep, staying asleep, or sleeping too much	0.83 (1.05)	1.70 (1.08)*	0.89 (1.09)	1.70 (1.06)*
PHQ 4	Feeling tired/having little energy	0.59 (0.76)	1.90 (1.21)*	0.64 (0.80)	1.96 (1.19)*
PHQ 5	Poor appetite or overeating	0.56 (0.83)	1.05 (1.05)	0.64 (0.95)	1.17 (1.07)
PHQ 6	Feeling bad about yourself/that you are a failure/letting people down	0.59 (0.88)	1.85 (0.99)*	0.55 (0.87)	2.00 (1.00)*
PHQ 7	Trouble concentrating	0.87 (0.95)	1.90 (1.17)*	0.87 (0.98)	1.83 (1.19)*
PHQ 8	Moving or speaking slowly/being fidgety or restless	0.31 (0.74)	0.60 (0.88)	0.40 (0.79)	0.65 (0.83)
PHQ 9	Thoughts that you would be better off dead/hurting yourself	0.20 (0.60)	1.30 (1.22)*	0.18 (0.54)	1.26 (1.14)*
Total PHQ-9 score		5.80 (4.29)	14.35 (5.22)*	6.10 (4.50)	14.74 (5.01)*
GAD-7		**15–19 Years old**		**Full sample**	
		Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Item #	Description	No diagnosis (n = 127)	Anxiety (n = 36)	No diagnosis (n = 245)	Anxiety (n = 45)
GAD 1	Feeling nervous, anxious, or on edge	0.51 (0.76)	1.33 (1.15)*	0.45 (0.71)	1.20 (1.14)*
GAD 2	Not able to stop or control worrying	0.62 (0.86)	1.39 (1.08)*	0.52 (0.80)	1.27 (1.07)*
GAD 3	Worrying too much about different things	0.58 (0.76)	1.31 (1.14)*	0.56 (0.79)	1.36 (1.17)*
GAD 4	Trouble relaxing	0.43 (0.74)	1.08 (1.20)*	0.44 (0.75)	1.11 (1.13)*
GAD 5	So restless that it is hard to sit still	0.55 (0.89)	0.94 (1.04)	0.57 (0.85)	0.87 (1.04)
GAD 6	Easily annoyed/irritable	1.08 (0.99)	2.19 (1.12)*	1.05 (1.04)	1.93 (1.21)*
GAD 7	Feeling afraid as if something awful might happen	0.53 (0.72)	1.47 (1.18)*	0.49 (0.70)	1.42 (1.18)*
Total GAD-7 score		4.31 (3.81)	9.72 (5.21)*	4.09 (3.68)	9.16 (5.08)*

GAD = Generalized Anxiety Disorder; PHQ = Patient Health Questionnaire; SD = standard deviation.
* $p < .05$, T-tests for mean differences for diagnosed and undiagnosed groups, with Bonferroni correction for multitesting (9 comparisons for depression items and seven comparisons for anxiety items).

population level. We provided estimates for adjusting population prevalence rates based on the tools' psychometric properties. Of note, high reported prevalence rates need substantial adjustment to prevent overestimation of the population burden.

Regarding cutoff scores that balance sensitivity and specificity, we identified a cutoff score of 10 or higher on the PHQ-9 to indicate a potential diagnosis of depression. Using this cutoff, the PHQ-9 demonstrated high sensitivity (91%) and good specificity (76%) for detecting depression among adolescents aged 10–19 years.

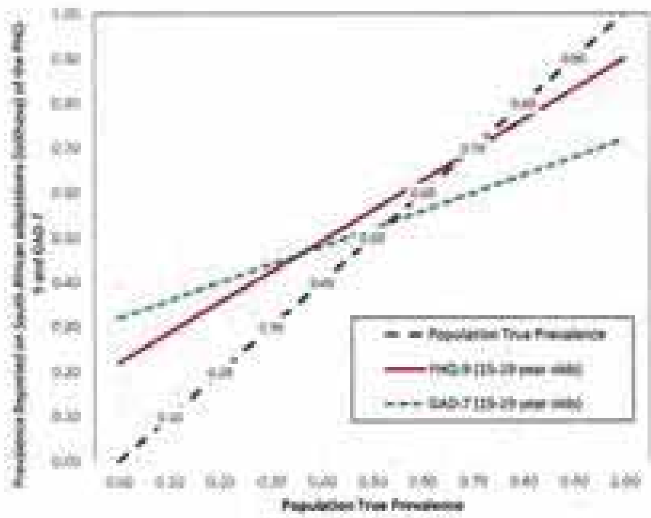
The GAD-7 demonstrated a 78% chance for discriminating between adolescents with and without an anxiety disorder. Cutoff scores with an optimal sensitivity-specificity balance were low (a score of six or more). Using this cutoff, the GAD-7 demonstrated moderate sensitivity (67%) and good specificity (75%). Limited validation research with adolescents is available to help us make sense of these findings. In Finland [23] and Ghana [8], the GAD-7 demonstrated factorial and construct validity in adolescents; however, neither study assessed criterion validity against a diagnostic assessment.

It is possible that the GAD-7 performed poorly compared to the PHQ-9 because of differences in the duration for these conditions in the K-SADS for a clinical diagnosis. The PHQ-9 and K-SADS use a 2-week period of symptoms. However, while the GAD-7 uses a 2-week period, the K-SADS requires 6 months of symptoms for a diagnosis of generalized anxiety disorder.

Information on individual item performance helps to identify items that do not discriminate between adolescents with and without a condition that could potentially be removed. Most items demonstrated a good discriminant ability, with the exception of "poor appetite or overeating" on the PHQ-9 and items related to movement on the PHQ-9 and GAD-7 ("*moving or speaking slowly/being fidgety or restless*" and "*so restless that it is hard to sit still*"). In two other LMICs (Nepal and Nigeria), items related to appetite also performed poorly in discriminating depression among adolescents [24]. It is possible that adolescents in general are more likely to struggle to sit still or to experience appetite fluctuations than adults, and therefore, these items did not work well to distinguish between adolescents with and without depression or anxiety when using the PHQ-9 or GAD-7. This notion however requires further exploration.

Strengths and limitations

The results should be considered in light of the following strengths and limitations. First, the adaptation and translation work conducted to produce the measures for validation was extensive and included local experts, adolescents, and their caregivers in the process. Another strength is the use of a comprehensive, high-quality diagnostic interview as the gold standard. In addition, we demonstrated that other cadres of mental health workers can be trained and supervised to conduct diagnostic assessments, an important finding in light of the shortages of mental health specialists in South Africa and many other LMICs.



	Population True Prevalence	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00
PHQ-9 (15-19 years old), Cut-off ≥11, Sensitivity=0.85, Specificity=0.78	Screening Tool Detected Prevalence	0.22	0.29	0.36	0.42	0.49	0.56	0.63	0.70	0.76	0.83	0.90
	Difference between true prevalence and detected prevalence	0.22	0.19	0.16	0.12	0.09	0.06	0.03	0.00	-0.04	-0.07	-0.10
	Positive Predictive Value (PPV)	0.00	0.31	0.51	0.64	0.73	0.80	0.86	0.91	0.94	0.97	1.00
	Negative Predictive Value (NPV)	1.00	0.99	0.97	0.95	0.92	0.89	0.84	0.77	0.66	0.46	0.00
GAD-7 (15-19 years old), Cut-off ≥6, Sensitivity=0.70, Specificity=0.68	Screening Tool Detected Prevalence	0.32	0.36	0.40	0.44	0.48	0.52	0.56	0.60	0.64	0.68	0.72
	Difference between true prevalence and detected prevalence	0.32	0.26	0.20	0.14	0.08	0.02	-0.04	-0.10	-0.16	-0.22	-0.28
	Positive Predictive Value (PPV)	0.00	0.22	0.35	0.49	0.60	0.69	0.77	0.84	0.89	0.95	1.00
	Negative Predictive Value (NPV)	1.00	0.96	0.91	0.85	0.78	0.71	0.62	0.51	0.38	0.21	0.00

Figure 2. Detected prevalence rates for anxiety and depression using the PHQ-9 and GAD-7 at different estimated true prevalence rates.

Because the study took place in one area in South Africa, using one language (isiXhosa), results may not be generalizable to other adolescent populations in South Africa or other countries. We used targeted recruitment to include adolescents likely to experience depression or anxiety, and our prevalence rates are possibly higher than what might be seen when conducting screening in the general population. Despite these efforts, we did not achieve the desired ratio between diagnosed and undiagnosed adolescents, leading to smaller sample sizes for analysis. Our study included an extremely small sample of younger adolescents with diagnoses. Future studies with larger samples of younger adolescents with depression and anxiety are needed to draw conclusions about the performance of the PHQ-9 and the GAD-7 in this age group.

We emphasized categorical analyses of disorders for the purpose of public health reporting, which typically requires categorical classification for reporting and policy-making. However, it is important to consider more continuous uses of these scales for tracking clinical improvements and for epidemiological research on risk and protective factors.

Conclusion

In global health, we cannot effectively manage what we do not measure [6]; therefore, addressing adolescent mental health starts with identifying appropriate tools to collect valid and reliable data on the prevalence of adolescent mental health conditions. These tools should enable both clinicians and researchers to engage with people in their home language and be adapted to the local context. In low-resource settings—both in research and practice—lines are often blurred between screening and diagnostic tools due to the lack of specialists and appropriate tools. Determining the validity of measurement tools and correctly interpreting that information is important to prevent tools from being used inappropriately. If we continue to use brief, self-report screening measures to make decisions about mental health service provision, it is important to confirm that these measures are valid and reliable in order to provide an accurate reflection of the mental health burden among adolescents and identify those at risk. This study found that the culturally adapted PHQ-9 tool had high sensitivity and specificity for both younger and older adolescents and could be used for population-level assessments of prevalence of depression among adolescents in the study setting. Further research is needed on adapting the GAD-7 to accurately capture the prevalence of generalized anxiety among adolescents in South Africa.

Acknowledgments

The authors are grateful to all the adolescents and their caregivers whose contributions made this study possible. The authors would like to acknowledge the hard work and perseverance of the Khayelitsha research team who conducted the study under exceptionally challenging circumstances: Athenkosi Mangele, Loyiso Ndamase, Luleka Sobekwa, Nonkululo Sikweyiya, Nozuko Matiso, Siphokazi Hlati, Vuyokazi Tasana, and Zwelebanzi Skiti, with the support of Zanele Siqabatiso and Zena Jacobs. Many members of this team were sick with COVID or lost family members due to COVID during the course of the study. The authors thank the various nongovernmental organizations and their staff for their partnership and support, both in referring adolescents to the study and for taking

on new referrals identified through the study. Their services are invaluable to the Khayelitsha community. The authors would like to thank Prof. Jason Bantjes from Stellenbosch University for his input and guidance in establishing the suicide referral protocols for the study. The authors are grateful to the Khayelitsha Day Hospital and team, who assisted with imminent suicide referral cases.

Funding Sources

This work was funded by the UNICEF HQ through the "Data Strengthening Grant" provided by the Bill & Melinda Gates Foundation in 2017–2018. This work was supported, in whole or in part, by the Bill & Melinda Gates Foundation [INV-001395]. Under the grant conditions of the Foundation, a Creative Commons Attribution 4.0 Generic License has already been assigned to the Author Accepted Manuscript version that might arise from this submission.

Supplementary Data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.jadohealth.2022.09.013.

References

[1] Solmi M, Radua J, Olivola M, et al. Age at onset of mental disorders worldwide: Large-scale meta-analysis of 192 epidemiological studies. Mol Psychiatry 2022;27:281–95.
[2] Patel V, Saxena S, Lund C, et al. The Lancet Commission on global mental health and sustainable development. Lancet 2018;392:1553–98.
[3] Erskine H, Baxter A, Patton G, et al. The global coverage of prevalence data for mental disorders in children and adolescents. Epidemiol Psychiatr Sci 2017;26:395–402.
[4] Jörns-Presentati A, Napp A-K, Dessauvagie AS, et al. The prevalence of mental health problems in sub-Saharan adolescents: A systematic review. PLoS One 2021;16:e0251689.
[5] Kohrt BA, Kaiser BN. Measuring mental health in humanitarian crises: A practitioner's guide to validity. Confl Health 2021;15:72.
[6] Hayes J, Carvajal L, Hijazi Z, et al. You can't manage what you do not measure – why adolescent mental health monitoring matters. J Adolesc Health 2021. https://doi.org/10.1016/j.jadohealth.2021.04.024.
[7] Carvajal L, Alsa JW, Requejo JH, et al. Measurement of mental health among adolescents at the population level: A multicountry protocol for adaptation and validation of mental health measures. J Adolesc Health 2021. https://doi.org/10.1016/j.jadohealth.2021.11.035.
[8] Adjorlolo S. Generalised anxiety disorder in adolescents in Ghana: Examination of the psychometric properties of the Generalised anxiety disorder-7 scale. Afr J Psychol Assess 2019;1:a10.
[9] Anum A, Adjorlolo S, Kugbey N. Depressive symptomatology in adolescents in Ghana: Examination of psychometric properties of the patient health questionnaire-9. J Affect Disord 2019;256:213–8.
[10] Buckley J, Otwombe K, Joyce C, et al. Mental health of adolescents in the era of antiretroviral therapy: Is there a difference between HIV-infected and uninfected youth in South Africa? J Adolesc Health 2020;67:76–83.
[11] Fatiregun AA, Kumapayi TE. Prevalence and correlates of depressive symptoms among in-school adolescents in a rural district in southwest Nigeria. J Adolesc 2014;37:197–203.
[12] Osborn TL, Venturo-Conerly KE, Wasil AR, et al. Depression and anxiety symptoms, social support, and demographic factors among Kenyan high school students. J Child Fam Stud 2020;29:1432–43.
[13] Smit W, de Lannoy A, Dover RV, et al. The built environment and non-communicable diseases in Khayelitsha, Cape Town. Health Place 2015;35:11–8.
[14] Cassidy T, Lloyd S, Bowman B. Violence, alcohol and symptoms of depression and in Cape Town's poorest communities: Results of a community survey. UCT working Paper. Available at: https://open.uct.ac.za/handle/11427/28271. Accessed October 23, 2022.
[15] Wolf EJ, Harrington KM, Clark SL, et al. Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. Educ Psych Meas 2013;76:913–34.
[16] Kroenke K, Spitzer RL, Williams JB. The PHQ-9: Validity of a brief depression severity measure. J Gen Intern Med 2001;16:606–13.
[17] Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing Generalized Anxiety Disorder: The GAD-7. Arch Intern Med 2006;166:1092–7.
[18] UNICEF. Measuring mental health among adolescents at the population level (MMAP): An Overview. New York: UNICEF; 2021. Available at: https://data.unicef.org/wp-content/uploads/2021/01/MMAP-Overview-January-2021.pdf. Accessed October 12, 2022.
[19] Kaufman J, Birmaher B, Brent D, et al. Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): Initial reliability and validity data. J Am Acad Child Adolesc Psychiatry 1997;36:980–8.
[20] American Psychiatric Association. Diagnostic and statistical manual of mental disorders: DSM-5. 5th ed. Washington: American Psychiatric Publishing; 2013.
[21] Mental health atlas 2020. Geneva: World Health Organization; 2021. Available at: https://www.who.int/publications/i/item/9789240036703. Accessed February 13, 2022.
[22] World Health Organization. Global Health Estimates: Life expectancy and leading causes of death and disability. Available at: https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates. Accessed January 28, 2022.
[23] Tiirikainen K, Haravuori H, Ranta K, et al. Psychometric properties of the 7-item Generalized Anxiety Disorder Scale (GAD-7) in a large representative sample of Finnish adolescents. Psychiatry Res 2019;272:30–5.
[24] Kohrt BA, Jordans MJ, Tol WA, et al. Validation of cross-cultural child mental health and psychosocial research instruments: Adapting the depression self-Rating scale and Child PTSD symptom scale in Nepal. BMC Psychiatry 2011;11:127.

International Journal of Mental Health and Addiction
https://doi.org/10.1007/s11469-025-01461-a

ORIGINAL ARTICLE

Meta-analysis of Reliability and Validity of the Bergen Social Media Addiction Scale (BSMAS)

Rossella Bottaro[1] · Mark D. Griffiths[2] · Palmira Faraci[1]

Accepted: 22 February 2025

Abstract
The present meta-analysis reviewed and summarized the psychometric properties of the
Bergen Social Media Addiction Scale (BSMAS), the most widely used tool for as
social media addiction (SMA) in research and clinical practice. Following the PI
2020 Statement guidelines, seven databases (*PubMed, PsycArticles, PsycInfo, Psy*
and Behavioral Sciences Collection, Medline, Wiley Online Library, and *Web of S*
were searched for studies reporting the dimensionality, item characteristics, reliabili
validity of the BSMAS. A total of 28 studies ($N = 62{,}406$) were reviewed. The uni
sionality of the BSMAS was unanimously confirmed with an optimal pooled Cro
alpha coefficient (0.83). Likewise, the pooled association between the BSMAS and a
depression, internet gaming disorder, and stress supported its construct validity. The
preliminary and encouraging evidence for other related measures and criteria, an
retest reliability, although these were qualitatively evaluated due to the limited nu
studies. Pending common nosographic categorization, the meta-analytic findings s
the appropriateness and validity of conclusions regarding SMA reached using the BS
Further evidence-based, randomized studies targeting various populations and sub
are warranted.

Keywords Bergen Social Media Addiction Scale · Meta-analysis · Psychometrics · Reliability · Validity

Bergen

Although not yet included in current official diagnostic manuals, social media addiction (SMA) is increasingly recognized as a mental health disorder. The growing interest by scholars is related to the proliferation of social networking sites (SNSs). There were 5.04 billion social profiles active in 2024, representing more than 62% of the world's population, who spend 2.23 hours a day using social media, mainly to stay in touch with friends and family (We Are Social & Meltwater, 2024).

✉ Rossella Bottaro
rossella.bottaro@unikorestudent.it; rossella.bottaro1@gmail.com

1 Psychometrics Laboratory, Department of Human and Social Sciences, University Kore of Enna, Cittadella Universitaria, 94100 Enna, Italy

2 International Gaming Research Unit, Psychology Department, Nottingham Trent University, Nottingham, UK

Published online: 08 March 2025

International Journal of Mental Health and Addiction
https://doi.org/10.1007/s11469-025-01461-x

ORIGINAL ARTICLE

Meta-analysis of Reliability and Validity of the Bergen Social Media Addiction Scale (BSMAS)

Rossella Bottaro[1] · Mark D. Griffiths[2] · Palmira Faraci[1]

Accepted: 22 February 2025

Abstract

The present meta-analysis reviewed and summarized the psychometric properties of the Bergen Social Media Addiction Scale (BSMAS), the most widely used tool for assessing social media addiction (SMA) in research and clinical practice. Following the PRISMA 2020 Statement guidelines, seven databases (*PubMed*, *PsycArticles*, *PsycInfo*, *Psychology and Behavioral Sciences Collection*, *Medline*, *Wiley Online Library*, and *Web of Science*) were searched for studies reporting the dimensionality, item characteristics, reliability, and validity of the BSMAS. A total of 28 studies ($N = 62,406$) were reviewed. The unidimensionality of the BSMAS was unanimously confirmed with an optimal pooled Cronbach's alpha coefficient (0.83). Likewise, the pooled association between the BSMAS and anxiety, depression, internet gaming disorder, and stress supported its construct validity. There was preliminary and encouraging evidence for other related measures and criteria, and test–retest reliability, although these were qualitatively evaluated due to the limited number of studies. Pending common nosographic categorization, the meta-analytic findings support the appropriateness and validity of conclusions regarding SMA reached using the BSMAS. Further evidence-based, randomized studies targeting various populations and subgroups are warranted.

Keywords Bergen Social Media Addiction Scale · Meta-analysis · Psychometrics · Reliability · Validity

Although not yet included in current official diagnostic manuals, social media addiction (SMA) is increasingly recognized as a mental health disorder. The growing interest by scholars is related to the proliferation of social networking sites (SNSs). There were 5.04 billion social profiles active in 2024, representing more than 62% of the world's population, who spend 2.23 hours a day using social media, mainly to stay in touch with friends and family (We Are Social & Meltwater, 2024).

Rossella Bottaro
rossella.bottaro@unikorestudent.it; rossella.bottaro1@gmail.com

1 Psychometrics Laboratory, Department of Human and Social Sciences, University Kore of Enna, Cittadella Universitaria, 94100 Enna, Italy

2 International Gaming Research Unit, Psychology Department, Nottingham Trent University, Nottingham, UK

Published online: 08 March 2025

SMA has been defined as being excessively concerned, overly preoccupied with social media, and spending considerable time and energy on social media that it interferes with an individual's ability to engage in social activities, form interpersonal relationships, study or work, or maintain health and wellbeing (Andreassen & Pallesen, 2014). However, a recent review of theories and models regarding SMA (Sun & Zhang, 2021) highlighted those terms such as "problematic social media use," "compulsive social media use," or "addictive social media use" which are often used interchangeably to refer to this type of behavior. Moreover, 25 different theories/models on the topic were identified, highlighting multiple comorbidities with psychiatric disorders such as anxiety and depression (Sun & Zhang, 2021; Szczygieł & Podwalski, 2020; Wang et al., 2022). Therefore, inconsistencies exist in the current literature regarding the definition and assessment of SMA. As a consequence, researchers and clinicians lack sufficient gold standards to guide their studies and practice.

The Bergen Social Media Addiction Scale (BSMAS; Andreassen et al., 2016) is one of the most commonly used psychometric instruments to assess SMA. The BSMAS is a six-item self-report scale with a five-point Likert scale response form ranging from *very rarely* (1) to *very often* (5) (e.g., "How often during the last year have you tried to cut down on the use of social media without success?") with a summed score ranging from 6 to 30. Each item reflects one of the six core criteria in the addiction components model of addiction (*i.e.*, salience, mood modification, tolerance, withdrawal, conflict, and relapse; Griffiths, 2005).

Notably, the BSMAS is an adaptation of the Bergen Facebook Addiction Scale (BFAS, Andreassen et al., 2012) also based on the six aforementioned criteria. In the adaption, the word "Facebook" was replaced with the words "social media," where social media is defined in the instructions to participants as "*Facebook, Twitter, Instagram, and the like.*" The one-factor structure of the BFAS was also adopted for the BSMAS. However, the scale's unidimensionality was assumed by the developers of the BSMAS although no confirmatory factor analysis (CFA) was reported in the study that first used it. Although this gap has since been filled by the numerous validations of the BSMAS in different languages, which has led to its widespread use, no meta-analytic evidence of its reliability and validity is currently available in the literature. Therefore, by systematically reviewing and meta-analyzing the published studies to date, the present study's main research questions (RQs) in relation to the psychometric studies that have been carried out internationally were (i) What are the characteristics of the studies that have psychometrically evaluated the BSMAS? (RQ1); (ii) What is the research quality of the studies that have evaluated the BSMAS? (RQ2); (iii) What is the consistency regarding the dimensionality of the BSMAS? (RQ3); How reliable is the BSMAS? (RQ4), (v) How valid is the BSMAS? and (vi) Is there any publication bias in the BSMAS studies?

Methods

A systematic review and meta-analysis of the previously published research on BSMAS was conducted in strict adherence to the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) guidelines (Page et al., 2021). It followed the three steps of identification, screening, and coding outlined below. The flow diagram is shown in Fig. 1, and the checklist can be found in Supplementary Table S1.

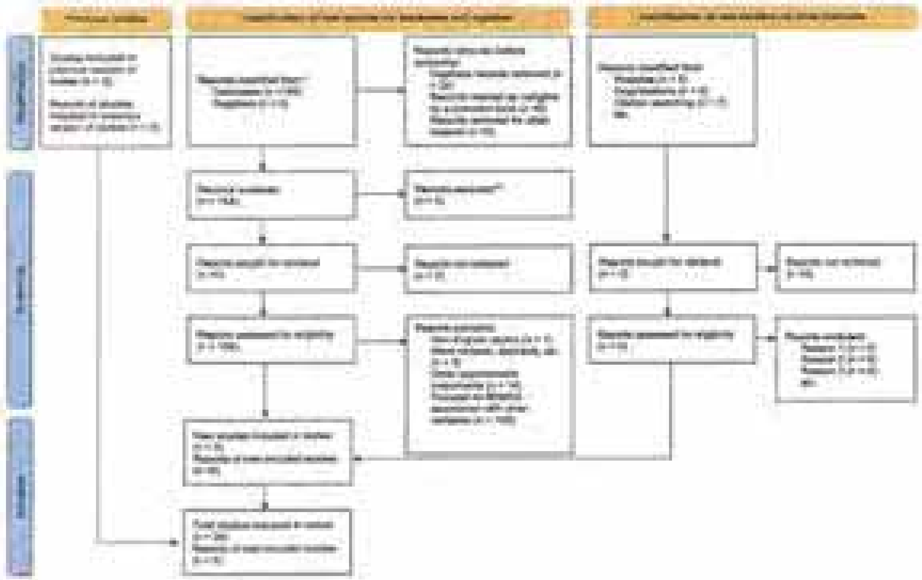


Fig. 1 PRISMA 2020 flow diagram the search strategy of the Bergen Social Media Addiction Scale (BSMAS) meta-analysis

Search Strategy

A comprehensive search of seven databases was conducted (i.e., *PubMed*, *PsycARTICLES*, *PsycINFO*, *Psychology*, and *Behavioral Sciences Collection*, *MEDLINE*, *Wiley Online Library*, and *Web of Science*). This search was supplemented through searching other sources (e.g., *Google Scholar*) and by manually searching the reference lists of the included studies. The authors carried out the search in August 2024. The years of the included studies ranged from 2016 (i.e., the year of the first publication using the BSMAS) to 2024. To identify the studies, the keyword "Bergen Social Media Addiction Scale" was searched for in paper titles and abstracts.

Screening

The following eligibility criteria were used to screen the selected studies: (i) being published in the English language, (ii) quantitative research published in peer-reviewed journals, and (iii) studies that specifically performed psychometric analysis of the BSMAS (e.g., dimensionality, reliability, and/or validity). The exclusion criteria were (i) previous literature reviews, books, theses, conference papers, and abstracts; (ii) studies that included BSMAS only as a measure of SMA in specific groups or models without examining its psychometric characteristics; (iii) studies that only provided the estimate prevalence of SMA using the BSMAS with the assumption that the reliability and validity of BSMAS had already been established; (iv) studies that only reported Cronbach's alpha to support the reliability of the scale in their study but did not specifically

focus on the psychometric properties of the BSMAS; and (v) studies that psychometrically evaluated the 18-item BSMAS.

The only exception to these exclusion criteria was the original study by the developers of the BSMAS (i.e., Andreassen et al., 2016). While it is not a validation study, given that it was the first study to use the BSMAS and it reported some psychometric properties, it was included due to its relevance to later studies.

Using the inclusion and exclusion criteria, one author first reviewed the titles and abstracts of the papers to determine which might be included. The authors used EndNote (X9; The EndNote Team, 2013) to initially search for the duplicate papers. The reference lists of the included papers were then searched for more potentially eligible studies. The initially selected papers were reviewed by a separate author. There were no differences in the inclusion or exclusion judgments made by the authors.

Coding

The included studies were coded for authors and year of publication, sample size, age of participants, BSMAS language, and online or offline administration procedure. Dimensionality and item properties were summarized by number of factors, method, and factor loadings. All available reliability and validity values were extracted. The authors chose Cronbach's alpha coefficient as the internal consistency index because it was present in all papers; any value ≥ 0.70 was considered acceptable. All Pearson's r coefficients were coded to summarize convergent, discriminant, and criterion validity with related measures.

Data Analyses

Firstly, the quality appraisal of the 28 included studies was assessed according to the Cochrane criteria (Higgins et al., 2023). It was plotted using the Risk of Bias VISualization (*robvis*; McGuinness & Higgins, 2021) tool for a generic study (Figs. 2 and 3). Moreover, the present study also used the COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) checklist (Mokkink et al., 2010) to evaluate the methodological quality of studies examining the psychometric properties of instruments. According to Schellingerhout et al. (2012), an overall score for the methodological quality of each included study for the main properties (i.e., 10 boxes) is obtained by taking the lowest rating of any item in a box (see http://www.cosmin.nl).

After coding the included studies, the information on dimensionality was presented qualitatively due to the absence of discrepancies between studies. Indeed, unidimensionality was unanimously supported despite the use of different methods and factor loadings. Test-retest and item-total correlations were also reported in narrative form when possible. Cronbach's alpha coefficients were chosen to summarize the reliability of the instrument in the present random-effect meta-analysis. For the longitudinal studies (Chen et al., 2020a, 2020b; Gomez et al., 2024; Shin, 2022), the baseline Cronbach's alpha was included. When studies reported two or more Cronbach's alpha values due to the inclusion of different subgroups, the mean value was used due to their similarity (e.g., 0.85 for Hong Kong and 0.82 for Taiwan; Leung et al., 2020).

Pearson's r correlation coefficient between BSMAS and other measures used were considered as reported in the included papers, as well as the effect size. This coefficient was then transformed into a Fisher-z score based on sample size (Lipsey & Wilson, 2001). According to the criteria formulated by Cohen (1988) for effect sizes, $r = 0.10$ was

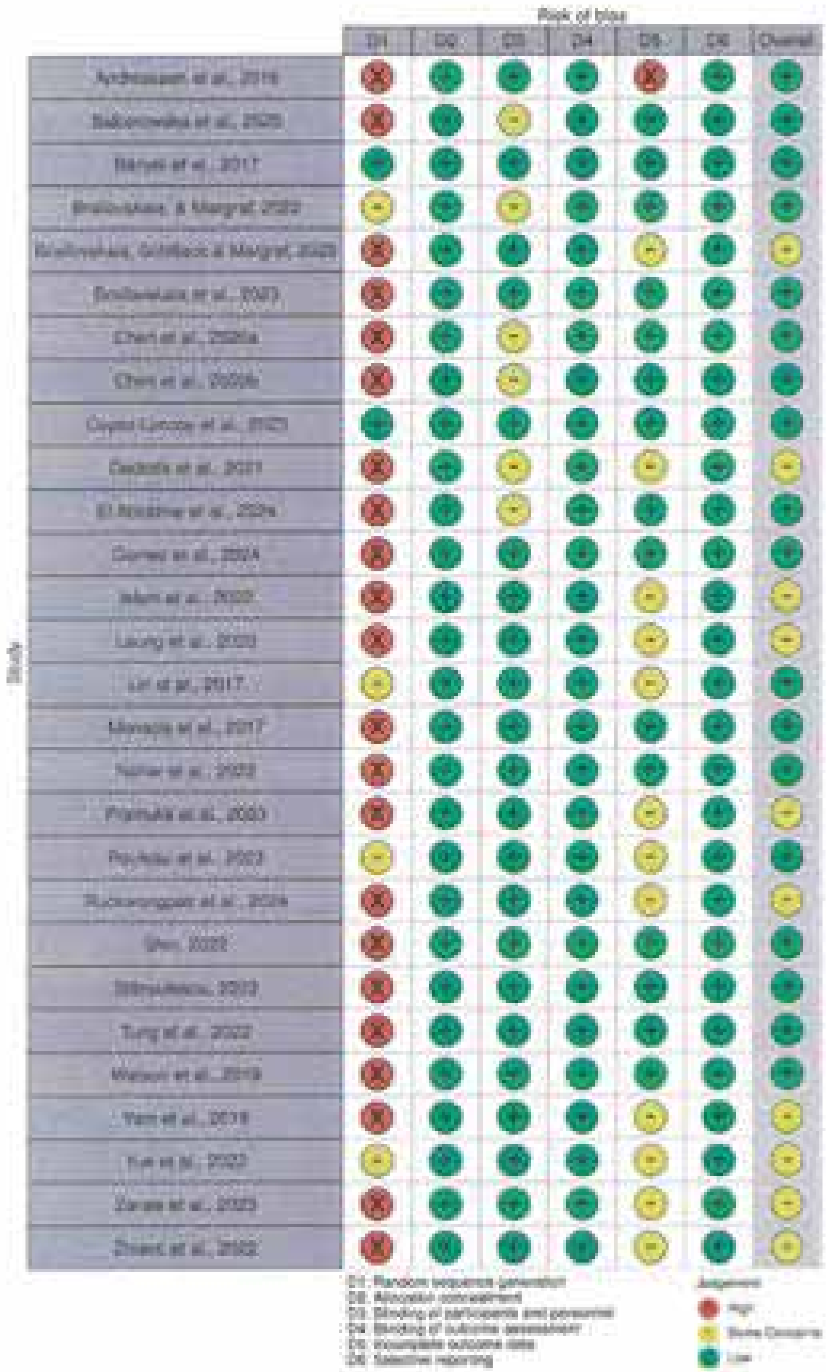


Fig. 3 Traffic-light plot of risk of bias assessment in the studies included in the meta-analysis

Fig. 1 Summary plot of risk of bias assessment

considered small, $r = 0.30$ medium, and $r = 0.50$ large. When the number of coefficients between studies was substantial (i.e., five or more studies), a quantitative meta-analysis was performed (e.g., when the association between SMA and anxiety was studied). However, when the association was examined in fewer than five studies, a qualitative synthesis was reported (Myung, 2023).

Cohen's Q statistics were used to evaluate the heterogeneity within the meta-analysis. The level of heterogeneity was assessed according to the Cochrane guidelines (Higgins et al., 2023), which were as follows: 0%–40% nonsignificant, 30%–60% moderate, 50%–90% substantial, and 75%–100% considerable. Based on the current available variables in the literature, moderation analyses were carried out to explain heterogeneity.

Publication bias was tested by performing Begg's and Egger's tests ($p < 0.1$, evidence of publication bias) and graphically using a funnel plot (Begg & Mazumdar, 1994; Egger et al., 1997). Moreover, the trim-and-fill method was performed to estimate the impact of potential missing studies due to publication bias in the funnel plot and to adjust the overall effect estimate (Duval & Tweedie, 2000a, 2000b; Shi & Lin, 2019). More specifically, the trim-and-fill method first *trims* the studies that cause a funnel plot's asymmetry and then *fills* the imputed missing studies in the funnel plot based on the bias-corrected overall estimate to minimize the impact of publication bias on the overall effect estimate. After assessing heterogeneity, publication bias was evaluated when the summary effect size was available.

Finally, data analysis was performed using the R software and programming language (R Core Team, 2021), using the "metafor" package for a random effects model (Viechtbauer, 2010). The 'rma.mv' function was employed for fitting linear (mixed-effects) models to metanalytic multivariate/multilevel fixed-effects and random/mixed-effects models, with or without moderators. Graphically, forest plots showed effect sizes and associated 95% confidence intervals for each study, and the funnel plot served as an estimate of publication bias.

Results

Sample Characteristics of Studies Evaluating the BSMAS (RQ1)

The initial search yielded 165 studies. After eliminating 22 duplicates, 154 papers were screened and 28 were selected based on the established eligibility criteria. The study

selection procedure, which was guided by the PRISMA Statement guidelines (Page et al., 2021), is shown in Fig. 1.

The psychometric properties of the Chinese BSMAS ($n=5$) and English BSMAS ($n=4$) were the most studied. However, 17 languages were used to validate the BSMAS in different countries, such as Italian, Hungarian, Greek, and Spanish (for more details, see Table 1), using a back-translation procedure. The studies comprised a total of 62,406 participants, with independent sample sizes ranging from 247 to 23,533, and a mean age of 21.89 years (SD = 4.25). Participants were mostly university students ($n=14$), adolescents ($n=7$), and adults ($n=7$).

Most of the included studies adopted a cross-sectional design except for three longitudinal studies (Table 1), were published in peer reviewed journals, and used the six-item version of the BSMAS (BSMAS mean score of all included studies: M = 14.99 [out of 30], SD = 4.52). Most data were collected through online surveys ($n=17$), others were conducted in an in-person setting ($n=8$) or both ($n=2$), and three studies did not specify this information. In total, 28 effect sizes were identified for Cronbach's alpha (ranging from 0.73 to 0.88). Moreover, internet gaming disorder, anxiety, depression, and stress were the most studied variables in evaluating construct validity.

Research Quality Assessment of BSMAS Studies (RQ2)

According to the Cochrane criteria for the classification of risk of bias, the included studies had a moderate overall risk (Figs. 2 and 3). Notably, the highest risk was related to the random sequence generation. The studies mostly used convenience samples, and surveys were distributed through social media platforms such as *Facebook*, *Instagram*, and *TikTok*, or internal university channels such as emails. The risk of incomplete outcome data raised concerns due to the lack of information about the mandatory response format (i.e., if the authors expected missing data) or as part of a larger research project.

Moreover, the authors assessed the quality of the included psychometric papers according to the COSMIN checklist (for details, see Supplementary Materials S1). As briefly shown in Table 2 and Fig. 4, all included studies provided optimal evidence of internal consistency by performing Cronbach's alpha ($\alpha \geq 0.70$). Most of them (26 studies out of 28) tested the construct validity of the BSMAS and evaluated its responsiveness. Moreover, the structural validity was well-tested in 22 of the 28 studies. Other studies found empirical support for the measurement invariance ($n=14$) and criterion validity ($n=11$) of the BSMAS. Finally, the results also highlighted that most of the included studies (22 studies out of 28) provided an adequate description of the BSMAS development process.

Dimensionality and Item Properties (RQ3)

All included studies supported the unidimensionality of the BSMAS by using the six-item version. The most widely used technique was confirmatory factor analysis (CFA; $n=24$), which showed satisfactory factor loadings on a single latent factor ($\lambda \geq 0.30$). Only one study performed an exploratory factor analysis (EFA; Naher et al., 2022) and two performed a principal component analysis (Brailovskaia et al., 2022; Žmavc et al., 2022). Satisfactory item-total correlations were found for all items (ranging from 0.49 for Item 4 to 0.73 for Item 2). Using an item response theory (IRT) approach, six studies (El Abiddine et al., 2024; Lin et al., 2017; Naher et al., 2022; Stănculescu, 2023; Tung et al., 2022; Zarate et al., 2023) provided further information about BSMAS items' characteristics.

Table 1 Studies characteristics ($N=28$)

Study	Language	N (% male)	Mean age ± SD	Target	Assessment setting	Method	Mean BSMAS ± SD	Data analysis	Reported goodness-of-fit indices	Factor loadings range	Cronbach's alpha
Andreassen et al. (2016)[a]	Norwegian	23,533 (35%)	35.8 ± 13.3	Adolescents and adults	Online	Cross-sectional	10.3 ± 4.77	NS	NS	NS	0.88
Balcerowska et al. (2020)	Polish	1099 (28.1%)	21.44 ± 2.85	Young adults	Online and offline	Cross-sectional	14.76 ± 4.71	CFA	$\chi^2=251.74$ CFI = 0.92, TLI = 0.87, RMSEA = 0.16 [0.14 0.17]	0.48–0.80	0.77
Bányai et al. (2017)	Hungarian	5961 (49.71%)	16.6 ± 0.94	Adolescents	Offline	Cross-sectional	NS	CFA + latent class analysis	$\chi^2=5836.19$ CFI = 0.95, TLI = 0.91, RMSEA = 0.07 [0.07 0.08]	0.60–0.81	0.85
Brailovskaia et al. (2022)	Chinese, France, German, Swedish, Polish, Russian, English, and Spanish	9418	NS	Adults	Online	Cross-sectional	NS	PCA	NS	NS	0.81 (China) 0.86 (France) 0.90 (Germany, Sweden) 0.89 (Poland) 0.87 (Russia, United Kingdom, United States) 0.85 (Spain)

Table 1 (continued)

Study	Language	N (% male)	Mean age ± SD	Target	Assessment setting	Method	Mean BSMAS ± SD	Data analysis	Reported goodness-of-fit indices	Factor loadings range	Cronbach's alpha
Brailovskaia, Schillack & Margraf (2020)	German	485 (22.3%)	24.75 ± 6.24	Adults	Online	Cross-sectional	NS	Point-biserial correlations and logistic regression analysis	NS	NS	0.82
Brailovskaia et al. (2023)	Lithuanian and German	2367 (43%)	30.28 ± 2.03	University students	Online and offline	Cross-sectional	13.89 ± 5.02	CFA + Latent class analysis	χ^2 = 28.62, CFI = 0.99, TLI = 0.99, RMSEA = 0.03 [0.02 0.05]	NS	0.81 (Lithuania) 0.83 (Germany)
Chen et al. (2020a)	Chinese	1108 (48.3%)	10.37 ± 0.95	Children	Offline	Cross-sectional	NS	CFA	χ^2 = 23.93, CFI = 0.99, TLI = 0.99, RMSEA = 0.04	0.59–0.73	0.73
Chen et al. (2020b)	Chinese	640 (41.56%)	22.34 ± 3.07	University students	Online	Longitudinal	NS	Time invariance	NS	NS	0.82
Copez-Lonzoy et al. (2023)	Spanish	650 (46.5%)	21.4 ± 2.7	University students	Online	Cross-sectional	12.78 ± 4.77	CFA	χ^2 = 23.93, CFI = 0.99, TLI = 0.99, RMSEA = 0.06	0.68–0.82	0.86
Dadiotis et al. (2021)	Greek	325 (18.2%)	21.6 ± 3.26	University students	NS	Cross-sectional	14.4 ± 4.37	CFA	NS	NS	0.75

Table 1 (continued)

Study	Language	N (% male)	Mean age ± SD	Target	Assessment setting	Method	Mean BSMAS ± SD	Data analysis	Reported goodness-of-fit indices	Factor loadings range	Cronbach's alpha
El Abiddine et al. (2024)	Arabic	757 (35.93%)	21.41 ± 2.87	University students	Offline	Cross-sectional	NS	CFA + Rash Analysis	CFI = 0.97, TLI = 0.94	0.40–0.67	0.74
Gomez et al. (2024)	English	276 (71%)	31.86 ± 9.94	Adults	Online	Longitudinal	20.79 ± 0.31	CFA + time invariance	χ^2 = 23.95 CFI = 0.96, TLI = 0.93, RMSEA = 0.08 [0.04 0.11]	0.66–0.80	0.88
Islam et al. (2022)	Bangladeshi	428 (90.89%)	16.13 ± 1.85	Adolescents	Offline	Cross-sectional	NS	CFA	χ^2 = 29.92 CFI = 0.97, TLI = 0.95, RMSEA = 0.08 [0.05 0.12]	0.56–0.76	0.86
Leung et al. (2020)	Chinese	642 (41.43%)	22.29 ± 3.14	Adults	Online	Cross-sectional	NS	CFA	(Hong Kong) χ^2 = 11.24 CFI = 0.99, TLI = 0.99, RMSEA = 0.03 [0.00 0.08] (Taiwan) χ^2 = 26.89 CFI = 0.97, TLI = 0.94, RMSEA = 0.08 [0.06 0.11]	0.61–0.72 (Hong Kong) 0.54–0.57 (Taiwan)	0.85 (Hong Kong) 0.82 (Taiwan)
Lin et al. (2017)	Persian	2676 (56.5%)	15.54 ± 1.21	Adolescents	NS	Cross-sectional	15.24 ± 4.83	CFA + Rash Analysis	CFI = 0.99, TLI = 0.99, RMSEA = 0.06	0.64–0.83	0.86

Table 1 (continued)

Study	Language	N (% male)	Mean age ± SD	Target	Assessment setting	Method	Mean BSMAS ± SD	Data analysis	Reported goodness-of-fit indices	Factor loadings range	Cronbach's alpha
Monacis et al. (2017)	Italian	769 (54%)	21.63 ± 3.93	Adolescents and young adults	Offline	Cross-sectional	14.2 ± 5.99	CFA + invariance across gender and age	χ^2 = 66.83 CFI = 0.98, RMSEA = 0.09 [0.07 0.11]	0.46–0.86	0.88
Naher et al. (2022)	Bangladeshi	577 (61.9%)	20.95 ± 1.92	University student	Offline	Cross-sectional	20.95 ± 1.92	EFA + CFA + IRT + Network Analysis	χ^2 = 6.62 CFI = 1.00, TLI = 1.00, RMSEA = 0.00	0.57–0.68	0.80
Pramukti et al. (2023)	Indonesian	458 (26%)	22.5 ± 8.07	University student	Online	Cross-sectional	16.68 ± 4.50	CFA	CFI = 1.00, TLI = 1.00, RMSEA = 0.00	0.54–0.78	0.80
Rouleau et al. (2023)	French	247 (23.9%)	14.70 ± 1.6	Adolescents	Online	Cross-sectional	NS	CFA	CFI = 0.98, TLI = 0.98, RMSEA = 0.07	0.56–0.80	0.84
Ruckwongpatr et al. (2024)	Thai	801 (33.1%)	20.69 ± 3.08	University students	Online	Cross-sectional	15.22 ± 4.69	CFA + invariance across gender	χ^2 = 19.13 CFI = 0.99, TLI = 0.98, RMSEA = 0.04 [0.01 0.06]	0.56–0.76	0.83
Shin (2022)	Korean	401 (27.2%)	21.9 ± 1.8	University students	Online	Longitudinal	11.88 ± 5.30	EFA + CFA	χ^2 = 19.78 CFI = 0.96, TLI = 0.97, RMSEA = 0.08 [0.03 0.12]	0.61–0.77	0.86

Table 1 (continued)

Study	Language	N (% male)	Mean age ± SD	Target	Assessment setting	Method	Mean BSMAS ± SD	Data analysis	Reported goodness-of-fit indices	Factor loadings range	Cronbach's alpha
Stănculescu (2023)	Romanian	705 (39%)	30.24 ± 9.15	Adults	Online	Cross-sectional	11.17 ± 5.47	CFA + IRT + Network Analysis	CFI = 0.98, TLI = 0.99, RMSEA = 0.06 [0.05 0.09]	0.74–0.94	0.84
Tung et al. (2022)	Malaysian	380 (23.9%)	24.04 ± 5.07	University students	Online	Cross-sectional	16.8 ± 5.40	CFA + Rash Analysis	χ^2 = 0.11 CFI = 0.99, TLI = 0.99, RMSEA = 0.04	0.68–0.76	0.86
Watson et al. (2019)	English	440 (49.1%)	17.3 ± 1.67	Adolescents	NS	Cross-sectional	16.02 ± U	CFA	χ^2 = 135.39 CFI = 0.82, TLI = 0.70, RMSEA = 0.18 [0.15 0.18]	0.45–0.73	0.77
Yam et al. (2019)	Chinese	307 (32.4%)	21.64 ± 8.11	University students	Online	Cross-sectional	NS	CFA	χ^2 = 17.06 CFI = 0.98, TLI = 0.96, RMSEA = 0.07	NS	0.82
Yue et al. (2022)	Mongolian	1120 (45.8%)	20.89 ± 1.4	University students	Offline	Cross-sectional	NS	CFA + invariance across gender	χ^2 = 39.02 CFI = 0.99, RMSEA = 0.05 [0.04 0.07]	0.63–0.73	0.82
Zarate et al. (2023)	English	968 (68.5%)	29.5 ± 9.36	University students	Online	Cross-sectional	NS	CFA + IRT	χ^2 = 13.35 CFI = 0.99, TLI = 0.99, RMSEA = 0.03 [0.00 0.06]	0.68–0.86	0.88

Table 1 (continued)

Study	Language	*N* (% male)	Mean age ± *SD*	Target	Assess-ment setting	Method	Mean BSMAS ± *SD*	Data analysis	Reported goodness-of-fit indices	Factor loadings range	Cronbach's alpha
Žmavc et al. (2022)	Slovenian	4868 (26.8%)	22.9 ± 3.19	Univer-sity stu-dents	Online	Cross-sec-tional	13.7 ± 5.75	PCA + CFA	CFI = 0.99, TLI = 0.99, RMSEA = 0.09	0.66–0.85	0.87

*original study; *NS* not specified, *CFA* confirmatory factor analysis, *EFA* exploratory factor analysis, *IRT* item response theory, *PCA* principal component analysis, *CFI* Comparative Fit Index, *TLI* Tucker-Lewis Index, *RMSEA* root mean square error of approximation

Springer

Table 2 COSMIN methodological quality of each study per measurement property and questionnaire

Study	PROM development	Content validity	Structural validity	Internal consistency	Cross cultural validation/measurement invariance	Reliability	Measurement error	Criterion validity	Hypotheses testing for construct validity	Responsiveness
Andreassen et al. (2016)*	Adequate	NA	NA	Very good	Very good	Inadequate	NA	Very good	Adequate	Very good
Balcerowska et al. (2020)	Doubtful	NA	Very good	Very good	NA	NA	NA	NA	Very good	Very good
Bányai et al. (2017)	Very good	NA	Very good	Very good	Very good	NA	NA	NA	Very good	Very good
Brailovskaia et al. (2022)	Very good	NA	Adequate	Very good	NA	NA	NA	NA	Very good	Very good
Brailovskaia, Schillack & Margraf (2020)	Very good	NA	NA	Very good	NA	NA	NA	NA	Very good	Very good
Brailovskaia et al. (2023)	Adequate	NA	Adequate	Very good	Very good	NA	NA	NA	Very good	Very good
Chen et al. (2020a)	Adequate	NA	Adequate	Very good	Very good	NA	NA	Very good	Very good	Very good
Chen et al. (2020b)	Adequate	NA	NA	Very good	Very good	NA	NA	Very good	Very good	Very good
Copez-Lonzoy et al. (2023)	Adequate	NA	Very good	Very good	Very good	NA	NA	NA	Very good	Very good
Dadiotis et al. (2021)	Adequate	NA	Very good	Very good	NA	NA	NA	Very good	Very good	Very good
El Abiddine et al. (2024)	Adequate	NA	Very good	Very good	Na	Na	Na	Na	Very good	Very good
Gomez et al. (2024)	Adequate	NA	Very good	Very good	Very good	Doubtful	Doubtful	NA	NA	NA

Table 2 (continued)

Study	PROM development	Content validity	Structural validity	Internal consistency	Cross cultural validation/measurement invariance	Reliability	Measurement error	Criterion validity	Hypotheses testing for construct validity	Responsiveness
Islam et al. (2022)	Doubtful	NA	Very good	Very good	NA	NA	NA	Very good	Very good	Very good
Leung et al. (2020)	Doubtful	NA	Very good	Very good	Very good	NA	NA	Very good	Very good	Very good
Lin et al. (2017)	Adequate	NA	Very good	Very good	Very good	NA	NA	Very good	Very good	Very good
Monacis et al. (2017)	Adequate	NA	Very good	Very good	Very good	NA	NA	NA	Very good	Very good
Naher et al. (2022)	Adequate	NA	Very good	Very good	Very good	NA	NA	Very good	Very good	Very good
Pramukti et al. (2023)	Adequate	NA	Very good	Very good	NA	NA	NA	NA	Very good	Very good
Rouleau et al. (2023)	Adequate	NA	Very good	Very good	Very good	NA	NA	NA	Very good	Very good
Ruckwongpatr et al. (2024)	Adequate	NA	Very good	Very good	Very good	NA	NA	Very good	Very good	Very good
Shin (2022)	Adequate	NA	Very good	Very good	NA	Inadequate	NA	NA	Very good	Very good
Stănculescu (2023)	Adequate	NA	Very good	Very good	NA	NA	NA	Very good	Very good	Very good
Tung et al. (2022)	Adequate	NA	Very good	Very good	NA	NA	NA	NA	Very good	Very good
Watson et al. (2019)	Adequate	NA	Very good	Very good	NA	NA	NA	NA	Very good	Very good
Yam et al. (2019)	Adequate	NA	Very good	Very good	NA	NA	NA	Very good	Very good	Very good

Springer

Table 2 (continued)

Study	PROM development	Content validity	Structural validity	Internal consistency	Cross cultural validation/measurement invariance	Reliability	Measurement error	Criterion validity	Hypotheses testing for construct validity	Responsiveness
Yue et al. (2022)	Adequate	NA	Very good	Very good	Very good	NA	NA	NA	Very good	Very good
Zarate et al. (2023)	Adequate	NA	Very good	Very good	NA	NA	NA	Very good	Very good	Very good
Žmavc et al. (2022)	Adequate	NA	Very good	Very good	NA	NA	NA	NA	Very good	Very good

NA not applicable, *PROM* Patient-reported Outcome Measurement

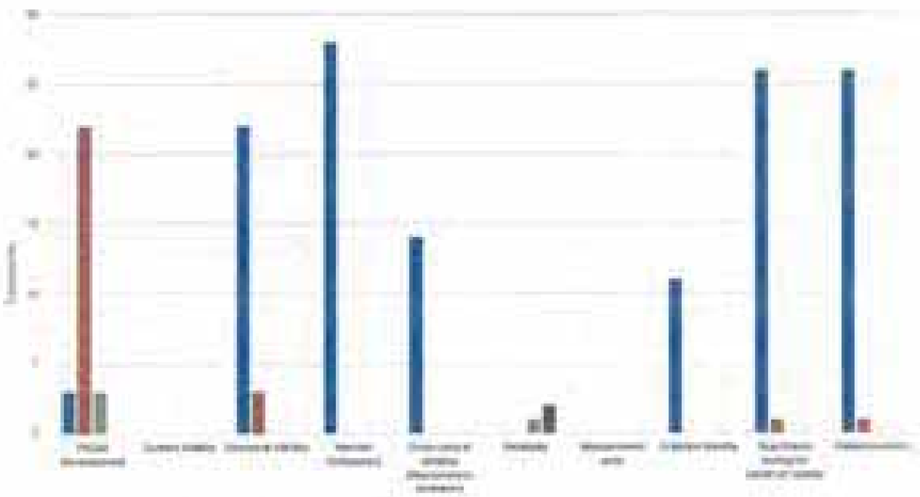


Fig. 4 Summarized methodological quality of each study per measurement property and questionnaire. *Note:* The "not applicable" label was excluded in the current representation; PROM = Patient reported Outcome Measurement

such as adequate discrimination and difficulty parameters for all items (Stănculescu, 2023; Tung et al., 2022), more information at middle and middle-high trait levels (peak ranged from +0.5 to +2.0 SD; Zarate et al., 2023), and no differential item function across gender or time spent on social media (Lin et al., 2017; Naher et al., 2022).

Reliability of the BSMAS (RQ4)

Internal consistency. All of the 28 included studies reported satisfactory Cronbach's alpha values, ranging from 0.73 to 0.88 (Fig. 5). Firstly, the Shapiro–Wilk's test suggested that the values were normally distributed ($p > 0.001$). After performing a random-effect meta-analysis, the estimated effect size was 0.83 (standard error = 0.01, $p < 0.001$, C.I. 95% [0.81–0.85]). However, the findings of the meta-analysis showed a substantial degree of heterogeneity ($I^2 = 98.53\%$, Cochran's $Q = 726.79$, $df = 26$, $p < 0.001$).

A moderator analysis of studies by sample size was performed. However, there were no differences between studies ($p = 0.07$) and considerable residual heterogeneity ($I^2 = 97.65\%$, Cochran's $Q = 447.62$, $df = 26$, $p < 0.001$). Subsequently, population type was also tested (i.e., university students or others) as a moderator which showed nonsignificant results ($p = 0.56$) with relevant residual heterogeneity ($I^2 = 98.42\%$, Cochran's $Q = 646.19$, $df = 26$, $p < 0.001$). Finally, the assessment setting (i.e., online vs. offline) was also tested as a moderator. However, there were no differences between the online and offline settings ($p = 0.49$), with noticeable residual heterogeneity ($I^2 = 98.39\%$, Cochran's $Q = 482.87$, $df = 22$, $p < 0.001$).

Test–retest reliability. Test–retest reliability was evaluated in only three studies. As a consequence of their small size and varying time-frames, the results here are presented qualitatively. Shin (2022) and Chen et al., (2020a, 2020b) showed strong associations at three-week and three-month intervals ($r = 0.75$ and 0.72, respectively). Similarly, Gomez

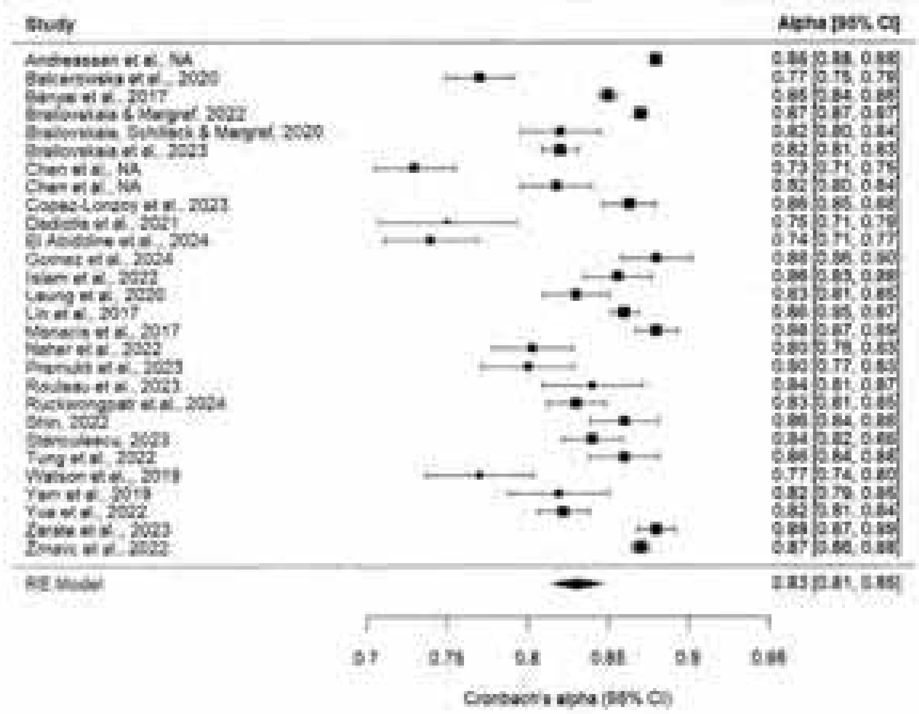


Fig. 5 Forest plot of overall Cronbach's alpha. *Note*: RE Model = random effect model; *original scale development study

et al. (2024) highlighted moderate reliability at 12 and 24 months (r = 0.53 and 0.42, respectively).

Validity of the BSMAS (RQ5)

Construct validity. The validity of the BSMAS has been studied in relation to relevant convergent and divergent constructs (forest plot in Fig. 6). For convergent validity, the association with depression (n = 11), anxiety (n = 11), internet gaming disorder (n = 9), and stress (n = 8) was widely investigated after verifying the assumption of normality (Shapiro–Wilk's test: p = 0.26, 0.15, and 0.24, respectively). The BSMAS association with anxiety showed medium effect size (r = 0.32) and high heterogeneity (I^2 = 96.85%, Cochran's Q = 302.53, df = 10, p < 0.001) with statistically significant results when sample size was tested as a moderator (F[1,9] = 2.35, p < 0.001). Similarly, the BSMAS association with depression showed medium effect size (r = 0.32) and high heterogeneity (I^2 = 96.23%, Cochran's Q = 518.86, df = 10, p < 0.001), with statistically significant results when sample size was tested as a moderator (F[1,9] = 214.73, p < 0.001). Moreover, the BSMAS association with internet gaming disorder showed a moderate-high effect size (r = 0.45) and high heterogeneity (I^2 = 99.69%, Cochran's Q = 3597.05, df = 8, p < 0.001) with significant results when sample size was tested as a moderator (F[1,7] = 2535.18, p < 0.001). Finally, the BSMAS association with stress showed a moderate effect size (r = 0.34) and high

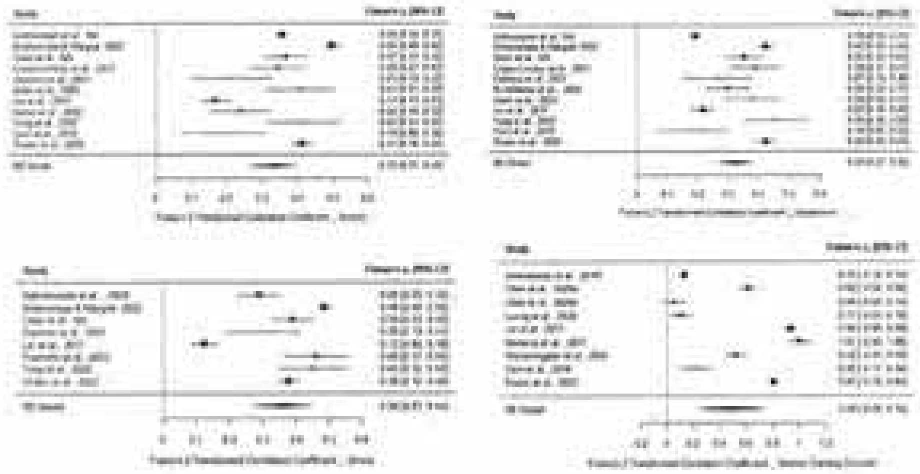


Fig. 6 Forest plot of overall effect size of construct validity measures. *Note*: Re Model = random effect model; * original study

heterogeneity (I^2 = 96.56%, Cochran's Q = 289.70, df = 7, p < 0.001) and was statistically significant when sample size was tested as a moderator (F[1,6] = 147.86, p < 0.001).

However, because they have not been thoroughly examined, a qualitative evaluation was undertaken when examining some variables. Negative correlations were found with self-esteem (ranging from −0.18 to −0.29; Dadiotis et al., 2021; Stănculescu, 2023; Žmavc et al., 2022), satisfaction with life (r = −0.23; El Abiddine et al., 2024), happiness (r = −0.11; Stănculescu, 2023), positive mental health and sense of control (r = −0.17 and −0.34; Brailovskaia et al., 2022), and conscientiousness, emotional stability, and general subjective well-being (r = −0.14, −0.23, and −0.15, respectively; Balcerowska et al., 2020). On the other hand, preliminary convergent validity results showed positive associations with *Facebook* addiction (r = 0.76 and 0.72; Balcerowska et al., 2020; Copez-Lonzoy et al., 2023), fear of missing out (r = 0.64; Copez-Lonzoy et al., 2023), loneliness (r = 0.12 and 0.40; Dadiotis et al., 2021; Naher et al., 2022), intensity of social media use (r = 0.54; Stănculescu, 2023), social anxiety (r = 0.21; Stănculescu, 2023), need to belong (r = 0.23; Stănculescu, 2023), anxious attachment (r = 0.43; Stănculescu, 2023), extraversion (r = 0.07; Balcerowska et al., 2020), social media engagement and addiction (r = 0.34; Copez-Lonzoy et al., 2023), nomophobia (r = 0.58; Pramukti et al., 2023), satisfaction with life (r = -0.23; El Abbidine et al., 2024), obsessive-compulsive disorder (r = 0.33; Andreassen et al., 2016), and attention-deficit/hyperactivity (r = 0.41; Andreassen et al., 2016). Lastly, only one study (Watson et al., 2020) examined the association between the BSMAS compared to other scales related to SMA such as Social Media Addiction Scale (Al-Menayes, 2015) and Social Media Disorder Scale (van den Eijnden et al., 2016). These findings showed significant associations (r = 0.54 and r = 0.44, respectively; Watson et al., 2020).

Criterion validity. Among the included studies, 11 highlighted criterion validity of BSMAS (Table 2). The most frequent criteria used to study the criterion validity of the BSMAS were time daily spent on the social media platform (n = 9) and time daily spent using the smartphone (n = 4). In most studies, participants were asked to report their daily use, but sometimes, the time frame was different and not comparable. In sum,

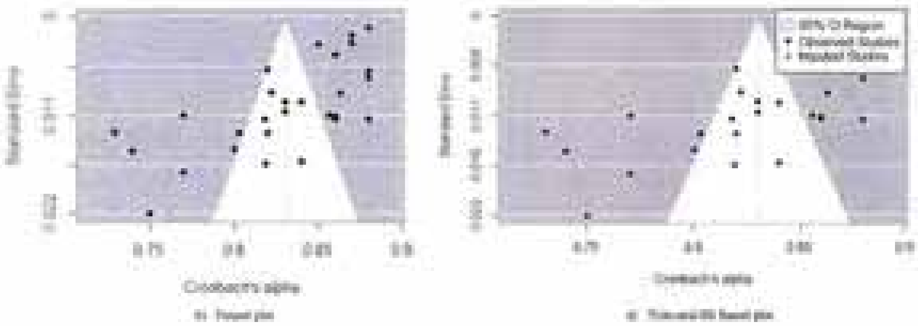


Fig. 7 Funnel plot of included studies' Cronbach's alpha

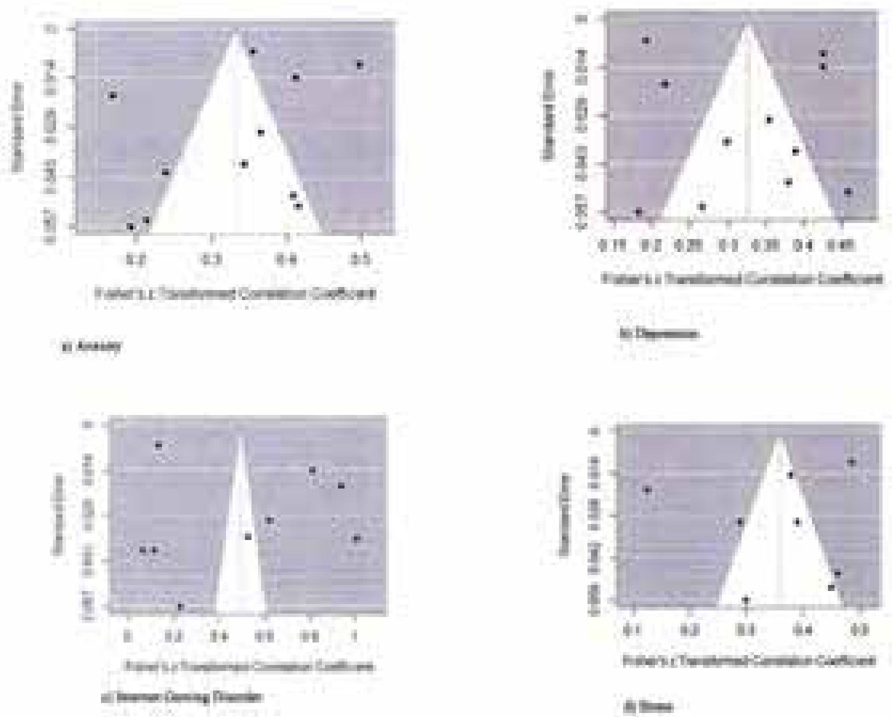


Fig. 8 Funnel plot of overall effect size of construct validity measures

the positive correlations were found between BSMAS and daily (i) social media use (ranging from 0.09 to 0.58; Brailovskaia & Margraf, 2022; Chen et al., 2020a, 2020b a, 2020b; Dadiotis et al., 2021; Lin et al., 2017; Naher et al., 2022; Yam et al., 2019; Žmavc et al., 2022) and (ii) smartphone use (ranging from 0.1 to 0.2; Chen et al., 2020a, 2020b; Yam et al., 2019). Only two studies reported the relationship with smartphone weekly use ($r=0.16$; Leung et al., 2020) and social media weekly use ($r=0.27$ and 0.1, respectively; Leung et al., 2020; Ruckwongpatr et al., 2024).

Publication Bias in BSMAS Studies (RQ6)

On a graphical level, a funnel plot was performed for Cronbach's alpha (Fig. 7a) and for anxiety, depression, internet gaming disorder, and stress (Fig. 8) with Fisher's z-transformed correlation coefficient used to estimate publication bias due to homogeneous data. The graphs showed symmetry and no results deviated significantly from the average effect size. This finding was also confirmed by Begg's test (Cronbach's alpha: Kendall's $\tau = -0.09$, $p = 0.76$; depression: Kendall's $\tau = -0.31$, $p = 0.02$; anxiety: Kendall's $\tau = -0.09$, $p = 0.76$; depression: Kendall's $\tau = -0.13$, $p = 0.65$; internet gaming disorder: Kendall's $\tau = -0.33$, $p = 0.26$; stress: Kendall's $\tau = -0.07$, $p = 0.90$) and Egger's test (Cronbach's alpha: $z = 1.81$, $p = 0.07$; anxiety: $z = 1.81$, $p = 0.07$; depression: $z = -0.98$, $p = 0.32$; internet gaming disorder: $z = -0.71$, $p = 0.47$; stress: $z = 0.69$, $p = 0.49$). Moreover, the estimation of publication bias was adjusted using the trim-and-fill method. The adjusted funnel plots are shown in Fig. 7b for Cronbach's alpha and in Fig. 9 for construct validity measures. More specifically, the trim-and-fill adjusted effect size was confirmed for the overall Cronbach's alpha (adjusted Cronbach's $\alpha = 0.33$), and for the construct validity measures with minimal differences (anxiety = 0.33, depression = 0.33, internet gaming disorder = 0.49, stress = 0.32).

Discussion

The present study aimed to meta-analyze the psychometric properties of the BSMAS, which is a widely used tool for assessing social media addiction (SMA). The systematic review identified 23 studies supporting its psychometric properties among 17 languages.

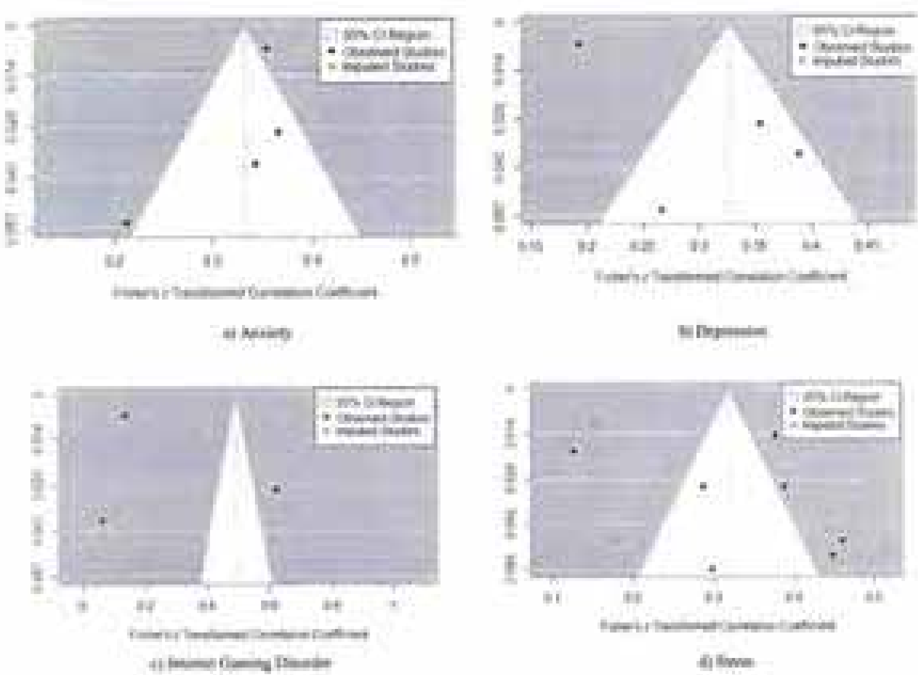


Fig. 9 Trim-and-fill adjusted funnel plot of overall effect size of construct validity measures

Therefore, the present study's findings significantly improve the current body of knowledge on the BSMAS. More specifically, the study is the first meta-analysis to simultaneously take into account dimensionality, internal consistency, test–retest reliability, and validity of all previously published studies on this topic. The COSMIN checklist suggested that the internal consistency, responsiveness, and construct validity of the BSMAS showed very good results based on the studies included in the meta-analysis. Although the association with other constructs has been analyzed, measurement invariance across different socio-demographic (e.g., gender, age, etc.) and cultural groups has not been well studied. Only a few studies have analyzed such aspects, and among individuals from different groups, making the results too inconsistent for further meta-analyses. Moreover, criterion validity was investigated in eleven studies examing different criteria (i.e., daily and weekly social media time and smartphone time). Future research is needed to differentiate between clinical and non-clinical populations due to the lack of shared diagnostic criteria. Further research is also needed for a gold standard in social media addiction diagnoses. Likewise, some psychometric features, such as content validity and measurement error, have yet to be conducted.

In relation to internal consistency, the findings showed that the Cronbach's alpha coefficient of the BSMAS across studies was good (0.84 in the random effect model). The findings overwhelmingly corroborated the unidimensionality of the six-item version of the BSMAS, mostly by performing CFA. However, other methods such as latent class analysis and item response theory have also supported the psychometric properties of the BSMAS. Although a few studies have used these methods, there are also other novel and innovative methodological analyses that could be carried out in future studies to assess the dimensionality of the BSMAS (e.g., bi-factor modeling). Moreover, the moderators examined were insufficient to account for the heterogeneity for reliability results. For instance, less than half of the included studies reported participants' time spent on smartphones or social media, and sometimes, this was reported in various time-frames, such as daily or weekly. Furthermore, assessment setting (i.e., online vs. offline), population type (i.e., university students vs. others), and sample size were not statistically significant moderators. This may have been because the possible categorizations were unbalanced. For example, there were 14 studies that collected data online with only six studies that collected the data offline. Although four moderators were tested, the source of heterogeneity remained unexplained. Further research should conduct other moderators analyses, for example, using cultural dimensions, socioeconomic factors, or psychometric differences in translation processes as moderators. The data available in the published papers did not provide sufficient information to test these hypotheses.

Although the preliminary results were encouraging, test–retest reliability was only investigated in three studies and at different time intervals. The currently available validity results demonstrated medium effect sizes ($r > 0.30$) for four relevant measures, namely, anxiety, depression, stress, and internet gaming disorder. In all cases, the high heterogeneity was explained by using sample size as a moderator. Although many studies are conducted in line with available resources, future research should include large samples for more accurate validation. Other associations with related constructs (e.g., loneliness, fear of missing out, and nomophobia were also examined, but the few results available did not allow for an overall evaluation. Moreover, time spent on smartphones or social media was shown to be an effective criterion for BSMAS criterion validity in only a few studies. Therefore, further research is needed to confirm the initial findings.

The results also showed low risk of publication bias. Additionally, these results were also confirmed after trim-and-fill adjustment. Indeed, the overall effect size variations

were negligible. Although a first inspection of the funnel plots might suggest that studies with unsatisfactory psychometric characteristics could not be published (e.g., Cronbach's alpha < 0.70), these results support the good psychometric properties of the BSMAS.

Finally, the present review also provided empirical evidence to support the use of the BSMAS to assess social media addiction. Given the current lack of a gold standard in psychiatric manuals such as the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5; America Psychiatric Association, 2013), the BSMAS could be used for theoretical reflection, clinical research and practice, and for potential clinical diagnosis and intervention within broader diagnostic frameworks as a valid tool with optimal psychometric properties across countries and settings. Luo et al. (2021) recently proposed a clinical cut-off for the BSMAS, but evidence is still limited and restricted to specific populations. Further research should particularly include clinical samples to achieve this goal. As suggested by the authors of the original version (i.e., Andreassen et al., 2016), who adapted the BSMAS from the BFAS (Andreassen et al., 2012), the BSMAS can be used to assess any addiction related to social media use, in contrast to other tools that focus on a specific social media platform such as *Facebook*, *Instagram*, and *YouTube*. This is an advantage considering how quickly new social media platforms can emerge and how quickly the popularity of some social media platforms can wane. However, other psychometric instruments to assess social media addiction are currently available and have been compared. For example, Watson et al. (2020) compared the BSMAS with the Social Media Disorder Scale (van den Eijnden et al., 2016) and the Social Media Addiction Scale (Al-Menayes, 2015), emphasizing the ability of the BSMAS to capture gender-specific differences in social media addiction. However, there are no psychometric reviews that have compared the currently available instruments. Although the present findings suggest the BSMAS is a candidate as a gold standard for the social media addiction assessment due to its good psychometric properties and established association with the well-being of users (Duradoni et al., 2020), future research should psychometrically compare other instruments to provide strong recommendations for clinical and research practice. According to Sigerson and Cheng's (2018) review of psychometric instruments, there are several promising scales to assess social media addiction, but ongoing psychometric research is crucial to provide a foundation of valid measurement in this timely research topic.

Limitations

The findings should be considered in light of some of the limitations. Firstly, the quality assessment of the included studies highlighted the extensive use of convenience samples in most studies, particularly university students, which reduces the generalizability of the results. Further psychometric evaluation of the BSMAS should comprise representative samples including older adults, clinical groups, and individuals from different cultural and educational backgrounds so that the psychometric properties of the BSMAS can be examined and compared among individuals from a wide variety of populations. In line with this concern, some studies used relatively small samples for their analysis ($N < 500$). Additionally, the online or offline assessment setting could be a significant source of bias due to poor standardization of procedures.

Notably, although the use of social media platforms to collect the data can be a limitation in terms of convenience sampling procedure, such a method of data collection can be an advantage for the specific purposes of the BSMAS because selecting participants

directly from online platforms allows for the confirmation of an essential selection criterion, namely, that the participants are social media users. Moreover, other limitations related to the search strategy, which did not include studies published in language other than English, such as the Turkish BSMAS which was published in its native language (i.e., Demirci, 2019), as well as a relatively small number of other studies, which may introduce selection bias. However, the highest quality journals and consequently the most important papers tend to be published in English. Future reviews carried out by large multi-lingual research teams could include psychometric research published in other languages. Findings from such studies may impact the findings of the present review by including from populations that are often little studied or with specific cultural peculiarities (e.g., poor or developing countries or countries where there is a significant digital divide).

Moreover, only three studies addressed test-retest reliability at varying time intervals. Therefore, the current evidence makes it difficult to support the stability of the results regarding the temporal reliability of the BSMAS. Further research should include longitudinal studies with standardized intervals to overcome this limitation. Likewise, the validity of BSMAS with respect to external criteria (e.g., time spent on social media, detrimental effects of excessive use) have been inconsistently operationalized across the published studies. Therefore, future studies should implement more standardized approaches to measuring these criteria because such inconsistencies could potentially affect the robustness of the psychometric findings. The currently available findings are so inconsistent that a comprehensive estimation of a coefficient was not possible. Globally, other studies are needed to investigate the psychometric properties of the BSMAS.

Lastly, the primary participants in the studies were mostly university students, and only one study (Rouleau et al., 2023) included psychiatric patients without including more specific information about their mental health. None of the studies included patients with or without a SMA diagnosis. This is evidently related to the fact that there is currently no consensus or gold standard for diagnosing SMA, such as in the fifth edition of the DSM-5 (American Psychiatric Association, 2013). As a consequence, the BSMAS's invariance across different groups, such as age, gender, or other target populations, should be enhanced in further research.

Conclusion

The present systematic review and meta-analysis highlighted encouraging psychometric properties of the BSMAS and its applications for further research. Evidence from 17 different languages confirmed its factor structure invariance and optimal internal consistency. Validity results supported the appropriateness and effectiveness of the inferences made using the instrument. In the near future, more evidence-based, randomized studies targeting various populations/subgroups are needed. Finally, the results could support the reflection on the definition of the nosographic criteria of SMA. In fact, due to the lack of a gold standard for quantifying and categorizing SMA, the use of a standardized method for SMA assessment would aid clinical research and practice in understanding individuals' suffering from this type of addiction.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11469-025-01461-x.

Data Availability Not applicable.

Declarations

Competing Interests The authors declare no competing interests.

References

Al-Menayes, J. (2015). Social media use, engagement and addiction as predictors of academic performance. *International Journal of Psychological Studies, 7*, 86–94. https://doi.org/10.5539/ijps.v7n4p86

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Publishing.

Andreassen, C. S., & Pallesen, S. (2014). Social network site addiction - An overview. *Current Pharmaceutical Design, 20*(25), 4053–4061. https://doi.org/10.2174/13816128113199990616

Andreassen, C. S., Torsheim, T., Brunborg, G. S., & Pallesen, S. (2012). Development of a Facebook addiction scale. *Psychological Reports, 110*(2), 501–517. https://doi.org/10.2466/02.09.18.PR0.110.2.501-517

Andreassen, C. S., Billieux, J., Griffiths, M. D., Kuss, D. J., Demetrovics, Z., Mazzoni, E., & Pallesen, S. (2016). The relationship between addictive use of social media and video games and symptoms of psychiatric disorders: A large-scale cross-sectional study. *Psychology of Addictive Behaviors, 30*(2), 252–262. https://doi.org/10.1037/adb0000160

Bányai, F., Zsila, Á., Király, O., Maraz, A., Elekes, Z., Griffiths, M. D., Andreassen, C. S., & Demetrovics, Z. (2017). Problematic social media use: Results from a large-scale nationally representative adolescent sample. *PLoS ONE, 12*(1), e0169839. https://doi.org/10.1371/journal.pone.0169839

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics, 50*(4), 1088–1101. https://doi.org/10.2307/2533446

Brailovskaia, J., & Margraf, J. (2022). Addictive social media use during Covid-19 outbreak: Validation of the Bergen Social Media Addiction Scale (BSMAS) and investigation of protective factors in nine countries. *Current Psychology, 43*, 13022–13040. https://doi.org/10.1007/s12144-022-03182-z

Brailovskaia, J., Schillack, H., & Margraf, J. (2020). Tell me why are you using social media (SM)! Relationship between reasons for use of SM, SM flow, daily stress, depression, anxiety, and addictive SM use – An exploratory investigation of young adults in Germany. *Computers in Human Behavior, 113*, 106511. https://doi.org/10.1016/j.chb.2020.106511

Brailovskaia, J., Truskauskaite-Kuneviciene, I., Kazlauskas, E., & Margraf, J. (2023). The patterns of problematic social media use (SMU) and their relationship with online flow, life satisfaction, depression, anxiety and stress symptoms in Lithuania and in Germany. *Current Psychology, 42*(5), 3713–3724. https://doi.org/10.1007/s12144-021-01711-w

Chen, I. H., Ahorsu, D. K., Pakpour, A. H., Griffiths, M. D., Lin, C. Y., & Chen, C. Y. (2020a). Psychometric properties of three simplified Chinese online-related addictive Behavior instruments among mainland Chinese primary school students. *Frontiers in Psychiatry, 11*, 875. https://doi.org/10.3389/fpsyt.2020.00875

Chen, I. H., Strong, C., Lin, Y. C., Tsai, M. C., Leung, H., Lin, C. Y., Pakpour, A. H., & Griffiths, M. D. (2020b). Time invariance of three ultra-brief internet-related instruments: Smartphone Application-Based Addiction Scale (SABAS), Bergen Social Media Addiction Scale (BSMAS), and the nine-item Internet Gaming Disorder Scale- Short Form (IGDS-SF9) (Study Part B). *Addictive Behaviors, 101*, 105960. https://doi.org/10.1016/j.addbeh.2019.04.018

Cohen, J. (1988). *Statistical power analysis for the behavioral Sciences*. Routledge.

Copez-Lonzoy, A., Vallejos-Flores, M., Capa-Luque, W., Salas-Blas, E., Montero Doig, A. M., Dias, P. C., & Bazo-Alvarez, J. C. (2023). Adaptation of the Bergen Social Media Addiction Scale (BSMAS) in Spanish. *Acta Psychologica, 241*, 104072. https://doi.org/10.1016/j.actpsy.2023.104072

Dadiotis, A., Bacopoulou, F., Kokka, I., Vlachakis, D., Chrousos, G. P., Darviri, C., & Roussos, P. (2021). Validation of the Greek version of the Bergen Social Media Addiction Scale in undergraduate students. *EMBnet J, 26*, e975. https://doi.org/10.14806/ej.26.1.975

Demirci, İ. (2019). Bergen Sosyal Medya Bağımlılığı Ölçeğinin Türkçeye uyarlanması, depresyon ve anksiyete belirtileriyle ilişkisinin değerlendirilmesi. *Anadolu Psikiyatri Dergisi, 20*, 15–22.

Duradoni, M., Innocenti, F., & Guazzini, A. (2020). Well-being and social media: A systematic review of Bergen addiction scales. *Future Internet, 12*(2), 24. https://doi.org/10.3390/fi12020024

Duval, S. J., & Tweedie, R. L. (2000a). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*(2), 455–463. https://doi.org/10.1111/j.0006-341x.2000.00455.x

Duval, S. J., & Tweedie, R. L. (2000b). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association, 95*(449), 89–98. https://doi.org/10.1080/01621459.2000.10473905

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629

El Abiddine, F. Z. E., Aljaberi, M. A., Alduais, A., Lin, C.-Y., Vally, Z., & Griffiths, D. M. (2024). The psychometric properties of the Arabic Bergen Social Media Addiction Scale. *International Journal of Mental Health and Addiction*. https://doi.org/10.1007/s11469-024-01297-x

Gomez, R., Zarate, D., Brown, T., Hein, K., & Stavropoulos, V. (2024). The Bergen-Social Media Addiction Scale (BSMAS): Longitudinal measurement invariance across a two-year interval. *Clinical Psychologist, 28*(2), 185–194. https://doi.org/10.1080/13284207.2024.2341816

Griffiths, M. (2005). A 'components' model of addiction within a biopsychosocial framework. *Journal of Substance Use, 10*(4), 191–197. https://doi.org/10.1080/14659890500114359

Higgins, J., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M., & Welch, V. (2023). Cochrane handbook for systematic reviews of interventions version 6.4. In Cochrane. www.training.cochrane.org/handbook

Islam, M. S., Jahan, I., Dewan, M. A. A., Pontes, H. M., Koly, K. N., Sikder, M. T., & Rahman, M. (2022). Psychometric properties of three online-related addictive behavior instruments among Bangladeshi school-going adolescents. *PLoS ONE, 17*(12), e0279062. https://doi.org/10.1371/journal.pone.0279062

Leung, H., Pakpour, A. H., Strong, C., Lin, Y. C., Tsai, M. C., Griffiths, M. D., Lin, Y. C., & Chen, I. H. (2020). Measurement invariance across young adults from Hong Kong and Taiwan among three internet-related addiction scales: Bergen Social Media Addiction Scale (BSMAS), Smartphone Application-Based Addiction Scale (SABAS), and Internet Gaming Disorder Scale-Short Form (IGDS-SF9) (Study Part A). *Addictive Behaviors, 101*, 105969. https://doi.org/10.1016/j.addbeh.2019.04.027

Lin, C. Y., Broström, A., Nilsen, P., Griffiths, M. D., & Pakpour, A. H. (2017). Psychometric validation of the Persian Bergen Social Media Addiction Scale using classic test theory and Rasch models. *Journal of Behavioral Addictions, 6*(4), 620–629. https://doi.org/10.1556/2006.6.2017.071

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. SAGE Publications, Inc.

Luo, T., Qin, L., Cheng, L., Wang, S., Zhu, Z., Xu, J., Chen, H., Liu, Q., Hu, M., Tong, J., Hao, W., Wei, B., & Liao, Y. (2021). Determination the cut-off point for the Bergen social media addiction (BSMAS): Diagnostic contribution of the six criteria of the components model of addiction for social media disorder. *Journal of Behavioral Addictions, 10*(2), 281–290. https://doi.org/10.1556/2006.2021.00025

McGuinness, L. A., & Higgins, J. P. T. (2021). Risk-of-bias VISualization (robvis): An R package and Shiny web app for visualizing risk-of-bias assessments. *Research Synthesis Methods, 12*(1), 55–61. https://doi.org/10.1002/jrsm.1411

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. (2010). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international Delphi study. *Quality of Life Research, 19*(4), 539–549. https://doi.org/10.1007/s11136-010-9606-8

Monacis, L., de Palo, V., Griffiths, M. D., & Sinatra, M. (2017). Social networking addiction, attachment style, and validation of the Italian version of the Bergen Social Media Addiction Scale. *Journal of Behavioral Addictions, 6*(2), 178–186. https://doi.org/10.1556/2006.6.2017.023

Myung, S. K. (2023). How to review and assess a systematic review and meta-analysis article: A methodological study (secondary publication). *Journal of Educational Evaluation for Health Professions, 20*, 24. https://doi.org/10.3352/jeehp.2023.20.24

Noher, L., Hiramoni, F. A., Alam, N., & Ahmed, O. (2022). Psychometric assessment of the Bangla version of the Bergen Social Media Addiction Scale. *Heliyon, 8*(7), e09929. https://doi.org/10.1016/j.heliyon.2022.e09929

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Bmj, 372*, n71. https://doi.org/10.1136/bmj.n71

Pramukti, I., Nurmala, I., Nadhiroh, S. R., Tung, S. E. H., Gan, W. Y., Siaw, Y. L., Yang, Y. N., Griffiths, M. D., Lin, C. J., & Pakpour, A. H. (2023). Problematic use of Internet among Indonesia university students: Psychometric evaluation of Bergen Social Media Addiction Scale and Internet Gaming Disorder Scale-Short Form. *Psychiatry Investigation, 20*(12), 1103–1111.

R Core Team (2021). *R: A language and environment for statistical computing*. In R Foundation for Statistical Computing. https://www.R-project.org/

Rouleau, R. D., Beauregard, C., & Beaudry, V. (2023). A rise in social media use in adolescents during the COVID-19 pandemic: The French validation of the Bergen Social Media Addiction Scale in a Canadian cohort. *BMC Psychology, 11*(1), 92. https://doi.org/10.1186/s40359-023-01141-2

Ruckwongpatr, K., Paratthakonkun, C., Sangtongdee, U., Pramukti, I., Nurmala, I., Angkasith, K., Thanachaisakul, W., Ketchatturat, J., Griffiths, M. D., Kao, K., & Lin, C. Y. (2024). Validity, reliability, and measurement invariance of the Thai Smartphone Application-Based Addiction Scale and Bergen Social Media Addiction Scale. *International Journal of Mental Health Promotion, 26*(4), 293–302. https://doi.org/10.32604/ijmhp.2024.047023

Schellingerhout, J. M., Verhagen, A. P., Heymans, M. W., Koes, B. W., de Vet, H. C., & Terwee, C. B. (2012). Measurement properties of disease-specific questionnaires in patients with neck pain: A systematic review. *Quality of Life Research, 21*(4), 659–670. https://doi.org/10.1007/s11136-011-9965-9

Shi, L., & Lin, L. (2019). The trim-and-fill method for publication bias: Practical guidelines and recommendations based on a large database of meta-analyses. *Medicine, 98*(23), e15987. https://doi.org/10.1097/MD.0000000000015987

Shin, N. Y. (2022). Psychometric properties of the Bergen Social Media Addiction Scale in Korean young adults. *Psychiatry Investigation, 19*(5), 356–361. https://doi.org/10.30773/pi.2021.0294

Sigerson, L., & Cheng, C. (2018). Scales for measuring user engagement with social network sites: A systematic review of psychometric properties. *Computers in Human Behavior, 83*, 87–105. https://doi.org/10.1016/j.chb.2018.01.023

Stănculescu, E. (2023). The Bergen Social Media Addiction Scale Validity in a Romanian sample using item response theory and network analysis. *International Journal of Mental Health and Addiction, 21*, 2475–2492. https://doi.org/10.1007/s11469-021-00732-7

Sun, Y., & Zhang, Y. (2021). A review of theories and models applied in studies of social media addiction and implications for future research. *Addictive Behaviors, 114*, 106699. https://doi.org/10.1016/j.addbeh.2020.106699

Szczygieł, K., & Podwalski, P. (2020). Comorbidity of social media addiction and other mental disorders – An overview. *Archives of Psychiatry and Psychotherapy, 22*(4), 7–11. https://doi.org/10.12740/APP/122487

The EndNote Team (2013). *EndNote*. In (Version EndNote X9) [64 bit]. Clarivate.

Tung, S. E. H., Gan, W. Y., Chen, J. S., Ruckwongpatr, K., Pramukti, I., Nadhiroh, S. R., Chang, Y. L., Lin, C. C., Pakpour, A. H., Lin, C. Y., & Griffiths, M. D. (2022). Internet-related instruments (Bergen Social Media Addiction Scale, Smartphone Application-Based Addiction Scale, Internet Gaming Disorder Scale-Short Form, and Nomophobia Questionnaire) and their associations with distress among Malaysian university students. *Healthcare, 10*(8), 1448. https://doi.org/10.3390/healthcare10081448

van den Eijnden, R. J. J. M., Lemmens, J. S., & Valkenburg, P. M. (2016). The Social Media Disorder Scale. *Computers in Human Behavior, 61*, 478–487. https://doi.org/10.1016/j.chb.2016.03.038

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Wang, Z., Yang, H., & Elhai, J. D. (2022). Are there gender differences in comorbidity symptoms networks of problematic social media use, anxiety and depression symptoms? Evidence from network analysis. *Personality and Individual Differences, 195*, 111705. https://doi.org/10.1016/j.paid.2022.111705

Watson, J. C., Prosek, E. A., & Giordano, A. L. (2020). Investigating psychometric properties of social media addiction measures among adolescents. *Journal of Counseling & Development, 98*(4), 458–466. https://doi.org/10.1002/jcad.12347

We Are Social, & Meltwater (2024). *Digital 2024 global overview report*. Retrieved August 21, 2024. From: https://datareportal.com/reports/digital-2024-global-overview-report"https://datareportal.com/reports/digital-2024-global-overview-report

Yam, C. W., Pakpour, A. H., Griffiths, M. D., Yau, W. Y., Lo, C. M., Ng, J. M. T., Lin, C. Y., & Leung, H. (2019). Psychometric testing of three Chinese online-related addictive behavior instruments among Hong Kong university students. *Psychiatric Quarterly, 90*(1), 117–128. https://doi.org/10.1007/s11126-018-9610-7

Yue, H., Zhang, X., Cheng, X., Liu, B., & Bao, H. (2022). Measurement invariance of the Bergen Social Media Addiction Scale across genders. *Frontiers in Psychology, 13*, 879259. https://doi.org/10.3389/fpsyg.2022.879259

Zarate, D., Hobson, B. A., March, E., Griffiths, M. D., & Stavropoulos, V. (2023). Psychometric properties of the Bergen Social Media Addiction Scale: An analysis using item response theory. *Addictive Behaviors Reports, 17*, 100473. https://doi.org/10.1016/j.abrep.2022.100473

Žmavc, M., Šorgo, A., Gabrovec, B., Crnkovič, N., Cesar, K., & Selak, Š. (2022). The protective role of resilience in the development of social media addiction in tertiary students and psychometric properties of the Slovenian Bergen Social Media Addiction Scale (BSMAS). *International Journal of Environmental Research and Public Health, 19*(20), 13178. https://doi.org/10.3390/ijerph192013178

11

EMPIRICAL ARTICLE

Do You "Like" My Photo? Facebook Use Maintains Eating Disorder Risk

Annalise G. Mabe, BS
K. Jean Forney, MS
Pamela K. Keel, PhD*

ABSTRACT

Objective: Social media sites, such as Facebook, merge two factors that influence risk for eating disorders: media and peers. Previous work has identified cross-sectional and temporal associations between Facebook use and disordered eating. This study sought to replicate and extend these findings using an experimental design.

Method: In Study 1, 960 women completed self-report surveys regarding Facebook use and disordered eating. In Study 2, 84 women were randomly assigned to use Facebook or to use an alternate internet site for 20 min.

Results: More frequent Facebook use was associated with greater disordered eating in a cross-sectional survey. Facebook use was associated with the maintenance of weight/shape concerns and state anxiety compared to an alternate internet activity.

Discussion: Facebook use may contribute to disordered eating by maintaining risk for eating pathology. As such, targeting Facebook use may be helpful in intervention and prevention programs. © 2014 Wiley Periodicals, Inc.

Keywords: eating disorders; social media; Facebook; body dissatisfaction; anxiety

(Int J Eat Disord 2014; 47:516–523)

Introduction

With 655 million daily users,[1] Facebook represents a ubiquitous merging of two social influences linked to risk for developing eating disorders through reinforcement of the thin ideal: media and peers (for recent review, see Keel and Forney[2]). Traditional media, such as movies, television, and magazines, portray an unrealistically thin ideal for female beauty.[3–5] Exposure to this ideal leaves many adolescent girls and women with body dissatisfaction,[6–8] which increases risk for disordered eating over time.[9,10] Peers influence risk for body dissatisfaction and eating pathology,[11–13] in part, by endorsing the thin ideal.[13] Today, college students use Facebook an average of 100 min/day, interacting with peers primarily by posting and viewing photos.[14] The ability to post carefully selected photos that may be digitally altered using online tools, such as "Plump & Skinny Booth,"[15] allows Facebook users to present and view images that adhere to unrealistic beauty ideals. Further, social media may reinforce the thin ideal by the posts, "likes," and comments of idealized images. Thus, it is important to understand whether and how the use of this common social media platform may influence risk for eating pathology.

Previous work has established small but significant associations between social media use and thin ideal internalization, body dissatisfaction, and eating pathology. Having a Facebook account was associated with greater thin ideal internalization, body surveillance, and drive for thinness in a large sample of adolescent girls.[16] Among those with Facebook accounts, number of "friends" and time spent on social media were significantly associated with increased body image disturbance.[16] Smith et al.[17] conducted a longitudinal study in college women in which they measured "maladaptive" Facebook use and changes in eating pathology over four weeks. Smith et al.[17] found that maladaptive Facebook use at baseline, defined as the tendency to seek out negative evaluations and/or engage in social comparisons, prospectively predicted greater eating pathology at follow-up. This effect was partially mediated by body dissatisfaction, suggesting that Facebook use may impact eating pathology via body dissatisfaction. Importantly, both maladaptive Facebook use and increases in disordered eating may be caused by an underlying third variable.

Accepted 13 January 2014

*Correspondence to: Pamela K. Keel, Department of Psychology, Florida State University, 1107 W. Call St., Tallahassee, FL 32306. E-mail: keel@psy.fsu.edu

Department of Psychology, Florida State University, Tallahassee, Florida

Published online 24 January 2014 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/eat.22254

Thus, an experimental design is needed to establish causation.

Study 1 aimed to replicate correlations between greater Facebook use and increased eating pathology.[16] Study 2 examined whether Facebook use causes temporal changes in eating disorder risk factors, specifically weight/shape concerns and anxiety,[18,19] and behavioral manifestation of concerns.[20,21] We hypothesized a positive correlation between higher Facebook use and higher disordered eating and that Facebook use would cause momentary increases in body dissatisfaction, anxiety, and urges to exercise. We focused on state anxiety because of its robust association with eating disorders.[22,23]

Study 1

Participants and Procedure

Nine hundred-sixty female college students completed a large screening instrument for a southeastern state university psychology subject pool in fall (n = 626) and spring (n = 334) semesters. Participants in the fall were significantly younger (M (SD) = 18.44 (.85) years) than in the spring (M (SD) = 19.10 (1.11) years), $t(958)$ = 10.34, p < .001, reflecting the passage of time. Across semesters, participants did not differ in ethnicity (18.45% Hispanic, $X^2(1)$ = .03, p = .87) or race (86.45% White, $X^2(3)$ = 2.14, p = .54). Participants received course credit. The university's institutional review board approved the study; informed consent was given prior to participation.

Measures

Eating Attitudes Test 26 (EAT-26)[24] assessed disordered eating attitudes and behaviors on a six-point scale from "Always" to "Never." The nonclinical scoring was used to ensure adequate sensitivity to individual differences.[25] Higher scores indicate greater disordered eating. The EAT-26 distinguishes between eating disorder cases and noncases,[26] and exhibits good convergent validity.[27] Due to limited space on the screening instrument, 19 items of the EAT-26 comprising the Dieting and Bulimia/Food Preoccupation subscales were used. Typical items include "I eat diet foods" and "I give too much time and thought to food." We used total scores from these subscales as a global measure of eating pathology, referred to as the EAT-19. Internal consistency of the EAT-19 was .92 in both the fall and the spring.

Duration of Facebook use was assessed with the question "How much time do you spend on Facebook per week?" Response options were 1 = "0 min," 2 = "<30 min," 3 = "30 min to < 1 h," 4 = "1 to <2 h," 5 = "2 to <4 h," 6 = "4–7 h," and 7 = ">7 h."

Results

The vast majority of women endorsed using Facebook on at least a weekly basis (97% in fall and 95.5% in spring). Mean (SD) scores for duration of Facebook use were 4.58 (1.52) in the fall and 4.74 (1.57) in the spring, reflecting approximately 2 h of Facebook use each week, with no significant difference in use between semesters, $t(958)$ = 1.53, p = .13. A small but significant positive correlation was observed between duration of Facebook use and disordered eating for participants in fall, $r(623)$ = .11, p < .01, and spring, $r(334)$ = .16, p < .01.

Study 2

Participants

Women (N = 84) included in Study 1 who endorsed Facebook use on a weekly basis (Facebook use ≥ 2) were recruited to participate in Study 2. Sociodemographic variables did not differ significantly between Study 1 and Study 2 participants. Participants identified as Caucasian (77.4%; n = 65), Hispanic (15.5%; n = 13), and African-American (7.1%, n = 6) and reported a mean (SD) age of 18.39 (.69) years. To ensure an adequate range of disordered eating, we used an enriched sampling design that balanced representation of individuals with low, medium, and high EAT-19 scores from Study 1 screens. Stratified randomization to the experimental and control groups was used to match disordered eating levels between conditions. EAT-26 scores from Study 2 did not differ between those randomly assigned to the experimental condition (M = 63.81; SD = 20.57) and the control condition (M = 63.25; SD = 18.81), $t(81)$ = −.13 , p = .90.

Procedure

After providing informed consent, participants completed a demographic survey, Visual Analog Scales (VASs; described below), and the State Trait Anxiety Inventory (STAI) State scale. Participants in the experimental group were instructed to log onto their Facebook account and spend 20 min on the site. Participants in the control group were instructed to use the internet for 20 min on Wikipedia researching the ocelot, a neutral rainforest animal, and on YouTube watching a preselected ocelot video. The control condition was designed to match the experimental condition on exposure to

text versus images while eliminating any images related to the human body. Participants were asked to remain on the assigned website(s) and not to use any links to connect to other sites to minimize risk of viral infections. This ensured that participants spent the entire 20 min in their assigned condition rather than connecting to another website. Participants were instructed to otherwise use the sites as they normally would. After 20 min of internet use, participants completed a second questionnaire packet, consisting of VAS, a STAI State scale, questions regarding their Facebook use, and the EAT-26 (described subsequently). Following internet use, the researcher cleared Internet browser history while the participant watched to ensure confidentiality of participants' personal information and compliance with study procedures. Upon completion, participants were debriefed and given class credit. The university's institutional review board approved this study.

Measures

Demographic information was collected with a brief survey that included questions about age, race, and ethnicity.

VAS ratings measured momentary experiences by having participants their level of "preoccupation with weight," "preoccupation with shape," and "urge to exercise," "RIGHT NOW" by placing a vertical line on a 100 mm horizontal line, anchored from "None at all" to "Extremely." The VAS is more sensitive to changes than Likert-type scale responses, as the latter may be influenced by recall of baseline answers.[28] Additional items probed for more serious disordered eating urges (e.g., "urge to vomit"). However, due to the small sample size and low base rate of these behaviors in a nonclinical sample, variance was too low to permit meaningful analyses. Preliminary analyses demonstrated significant robust correlations between responses on the VAS scales of "preoccupation with weight" and "preoccupation with shape" (Time 1 $r(75) = .89$, $p < .001$; Time 2 $r(75) = .95$, $p < .001$). Thus, these VAS items were averaged into a single "preoccupation with weight/shape" variable at Time 1 and Time 2 for analyses. VAS have successfully been used in other experimental studies examining changes in mood and body image over similar time frames.[29–31]

Eating Attitudes Test. In Study 2, participants completed the full EAT-26.[24] Test–retest reliability from the 19 items administered in Study 1 and Study 2 was high, $r(83) = .90$, $p < .001$. Cronbach's alpha for the EAT-26 was .91 in Study 2.

State Trait Anxiety Inventory. The STAI State subscale measured current anxiety before and after internet use.[32] This questionnaire assesses responses to questions such as "I feel nervous" on a four-point scale ranging from "almost never" to "almost always." Internal consistency was high (Time 1 $\alpha = .92$; Time 2 $\alpha = .93$).

Facebook survey questions were developed to understand the amount of time spent using Facebook, participants' activities on Facebook (e.g., viewing photos of friends, posting updates), importance of Facebook features (e.g., receiving comments or "likes" on their photos and posts), and access to Facebook (e.g., via a smartphone). Survey items are included in the Appendix. For participants assigned to the Facebook condition, an additional question evaluated how similar the 20 min of use was to their typical use of Facebook, with responses on a five-point scale ranging from "Not at all" to "Completely." Participants in the experimental condition indicated that their Facebook use was "Moderately" to "Very" representative of their typical use ($M = 3.52$; $SD = 1.19$). To evaluate how participants used Facebook, items were analyzed individually, and a Facebook score was created from Items 9, 10, 11, 12, 13, 14, 16, and 17. Facebook score reflects the importance and frequency of using Facebook features posited to heighten weight/shape concerns. Internal consistency of the Facebook score was good, $\alpha = .85$.

Data Analyses

Correlations examined the association between disordered eating and both Facebook items and Facebook score. Repeated measures analysis of variance assessed the effect of experimentally manipulated Facebook use as a between-subjects variable on within-subject changes in momentary ratings of "preoccupation with weight/shape," state anxiety, and "urge to exercise." Significant group × time interaction effects were followed by post hoc comparisons.

Results

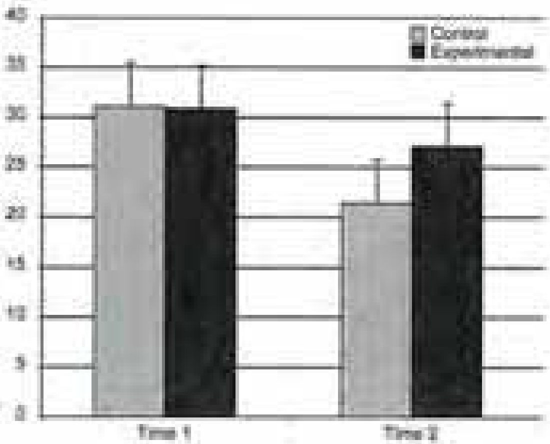
In Study 2, a similar effect size was found for the association between time spent on Facebook and EAT-26 score; however, due to the smaller sample size of Study 2, this association was not statistically significant, $r(83) = .09$, $p = .44$. Among the 84 participants in Study 2, mean (SD) time reported per Facebook session was 20.06 (17.75) min, while mean (SD) total time per day on the site was 76.28 (68.70) min. Most participants used Facebook daily ($M = 6.46$; $SD = 1.22$). The majority (91.7%; $n =$

TABLE 1. Effect of Facebook use on momentary disordered eating attitudes and feelings

VAS Scale	Group Mean (SE)	Pre Mean (SE)	Post Mean (SE)	Group F(1,73)	Time F(1,73)	Group × Time F(1,73)
Weight/shape preoccupation		31.81(2.97)	24.27 (3.61)	.22	23.62[a]	4.52[b]
Control	26.28 (4.14)	31.12 (4.25)	21.43 (4.29)			
Experimental	29.60 (4.09)	30.90 (4.17)	27.11 (4.23)			
STAI state anxiety[c]		34.99 (1.09)	34.87 (1.20)	1.50	0.05	7.55[d]
Control	36.29 (1.39)	37.08 (1.56)	35.50 (1.72)			
Experimental	33.57 (1.53)	32.90 (1.52)	34.24 (1.67)			
Urge to exercise		36.42 (3.33)	31.85 (3.27)	0.55	11.48[a]	0.002
Control	31.72 (4.63)	34.03 (4.75)	29.41 (4.66)			
Experimental	36.55 (4.57)	38.82 (4.68)	34.29 (4.59)			

[a] $p < .001$.
[b] $p < .05$.
[c] $df = 1, 72$.
[d] $p < .01$.

FIGURE 1 Visual analogue scale ratings of weight/shape concerns before and after internet use. Error bars represented standard errors of the mean.



77) of participants endorsed having a smart phone; 94.8% (n = 73) of those with a smartphone endorsed using a Facebook application. Among Facebook activities, 66.7% (n = 56) answered that they choose to look at photos over other activities.

EAT-26 scores were significantly associated with scores on several of the individual Facebook items used to create the Facebook score. Participants with greater disordered eating endorsed greater importance of receiving comments on their status ($r(83)$ = .32, $p < .01$) and photos ($r(83)$ = .29, p = .01), and greater importance of receiving "likes" on their status ($r(83)$ = .29, p <.01). Those with greater eating pathology reported untagging photos of themselves more often ($r(83)$ = .34, p = < .01) and endorsed comparing their photos to their female friends' photos more often ($r(83)$ = .22, p = .04). Disordered eating was not associated with the importance of "likes" on photos ($r(83)$ = .16, p = .16), nor with the frequency of changing profile pictures ($r(83)$ = −.15, p = .18). Consistent with study hypotheses, those with the greatest disordered eating had higher Facebook scores, $r(83)$ = .38, $p < .001$.

To understand the causal effects of Facebook use, comparisons were made between the experimental and control condition over time in weight/shape concerns, state anxiety, and urge to exercise (see **Table 1**). Participants in both conditions endorsed a decrease in their preoccupation with weight and shape from immediately before to immediately after spending 20 min on the internet, $p < .001$, d = .30. A significant group by time interaction (p = .04) indicates that the effect of time depended upon condition. Specifically, participants in the control group demonstrated a greater decline in weight/shape preoccupation than did participants who spent 20 min on Facebook. Post hoc comparisons supported a significant decrease in weight/shape preoccupation in controls ($F(1,36)$ = 21.29, $p < .001$, d = .42) and a less robust decline in experimental participants ($F(1,37)$ = 4.34, p = .04, d = .13). Weight/shape preoccupation did not significantly differ between conditions at Time 1 ($F(1,73)$ = .001, p = .97, d = .01) or Time 2 ($F(1,73)$ = .86, p = .35, d = −.22). The significant interaction effect remained after controlling for EAT-26 scores, suggesting that Facebook use maintains a preoccupation with weight and shape compared to an internet control condition (see **Fig. 1**).

Across conditions, state anxiety was maintained over time (p = .82). However, the effect of time varied by condition ($p < .01$). Specifically, participants in the control condition endorsed a significant decrease in anxiety ($F(1,35)$ = 6.04, p = .02, d = .56) while participants in the experimental condition endorsed a nonsignificant increase in anxiety ($F(1,37)$ = 2.57, p = .12, d = −.13). Post hoc comparisons supported no significant differences between experimental and control participants at Time 1 ($F(1,72)$ = 3.72, p = .06, d = .44) or at Time 2 ($F(1,72)$ = .28, p = .60, d = .12). The significant interaction effect, which remained once controlling

for EAT-26 scores, suggests that Facebook use maintains state anxiety compared to an alternative internet activity.

Urge to exercise decreased after spending 20 min on the Internet ($p < .001$, $d = .26$). The effect of time did not depend on condition ($p = .46$), suggesting that general internet use, not Facebook use specifically, is associated with the decrease.

Discussion

Before the advent of social media sites, women were confronted with unrealistically thin images of beauty from magazines, films, and television. Women also engaged with peers who represented a full range of expected body weights and shapes in their immediate environment, but could reinforce the thin ideal through discussions and behaviors. Now, women have a constant and active space to engage in social comparison with peers who may simultaneously portray and reinforce the thin ideal. Replicating previous research,[16] we found a significant but small association between Facebook use and disordered eating levels in two large samples of college-aged women. In addition, how women use Facebook (reflected by higher Facebook score) was associated with greater disordered eating. While previous longitudinal findings[17] reinforce that maladaptive patterns of Facebook use precede increases in disordered eating, our experimental design indicates that typical Facebook use may contribute to maintenance of weight/shape concerns and state anxiety, both of which are established eating disorder risk factors.[18,19] To the extent that these effects could be discerned after only 20 min of typical Facebook use in a laboratory setting raises concerns about how use of the site throughout the day may impact eating disorder risk.

Of interest, in our experimental design, internet use, regardless of condition, was associated with decreases in weight and shape preoccupation and urge to exercise. Such state changes may negatively reinforce internet use, explaining the widespread use of the internet for entertainment. Facebook users may not be aware of this cost because the overall experience may be positive. Without a non-internet control condition, it is unclear whether the observed main effects are specifically due to internet use or the passage of time more generally. Therefore, future research should seek to replicate these effects using a noninternet control condition.

Women with greater eating pathology not only reported spending more time on Facebook in Study 1, but also reported engaging in appearance-focused behaviors, such as comparing their appearance to friends' pictures and untagging photographs of themselves, perhaps in order to remove unflattering photographs and minimize opportunities to become the target of downward social comparison. In line with self-reported behaviors on Facebook, those who placed greater importance on the responses elicited by their Facebook content reported greater eating pathology. Specific aspects of use (e.g., social comparison to photos of peers) should be examined as potential mediators of the relationship between Facebook use and the maintenance of eating disorder risk. Alternatively, tendencies toward social comparison may serve as a moderator of the influence of Facebook use on eating disorder risk. Replication in larger samples would help to untangle these potential associations between individual and social risk factors.

Pending replication of these and other findings,[17] Facebook could be targeted as a maintenance factor in prevention programs. For example, interventions could address the implications of appearance-focused comments such as "you look so thin" or "I wish I had your abs," in perpetuating the thin ideal on Facebook, much as "fat talk" perpetuates this ideal in everyday conversations. An adaption of the "Fat Talk Free" campaign[33] as well as adaptations of media literacy programs[34,35] could encourage girls and women in the responsible use of social media sites. Similarly, if research finds that photoenhancing technology is common, advocacy may be effective in reducing the use of photoenhancing technology to promote unrealistic ideals.

The current studies benefited from a large college sample and measures with good psychometric properties. Our sampling approach ensured a range of disordered eating levels, allowing greater generalizability. Participants were representative of other college samples studied, as evidenced by comparable estimates of reported time on Facebook.[14] However, results should be interpreted with limitations in mind. We cannot rule out the possibility that our results reflect demand characteristics. Importantly, such effects should have prevented our observation that Facebook use was associated with decreases in both weight/shape concerns and urges to exercise, suggesting that changes (or the lack thereof) did not represent participants' efforts to unconsciously support our hypotheses. However, the opposite may be true; our findings may underestimate the effect of Facebook use on maintaining weight/shape concerns. Our control condition allows inferences about Facebook use compared to one other internet activity and may not generalize to other activities.

Specifically, the interaction effects observed in this study were driven, in part, by an observed decrease in weight and shape concerns in the control condition. This decrease over time may not be observed in naturalistic environments. An ecological momentary design[36] may better capture natural changes in affect as well as weight and shape preoccupation in relation to Facebook use. Our study does not address whether Facebook use influences eating disorder risk above and beyond other social or media influences. Future research should compare face-to-face social interaction to Facebook use. Additionally, the use of interviews about eating and Facebook use in future research would enhance understanding of the observed associations. As we measured momentary changes in risk factors, our results do not address whether Facebook use may contribute to actual eating disorders. However, the maintenance of risk is important to identify for prevention efforts.

Advances in technology may be impacting the nature of risk factors for disordered eating pathology in women. While the overall use of Facebook has a small but significant association with disordered eating, specific aspects of use demonstrate more robust associations with disordered eating. In addition, we found evidence that Facebook use may maintain preoccupation of weight and shape and state anxiety, both well-replicated risk factors for eating pathology. Future longitudinal research using ecological momentary assessment[36] in both at-risk and eating disordered populations would allow better understanding of the effects of Facebook use over time in a natural setting. As technology continues to change, more research is needed to understand the effects of social media in maintaining risk for eating disorders and other psychological problems.

References

1. Facebook. Key Facts. 2013. Available at: http://newsroom.fb.com/Key-Facts. Accessed June 11, 2013.
2. Keel PK, Forney KJ. Psychosocial risk factors for eating disorders. Int J Eat Disord 2013;46(5):433–439.
3. Polivy J, Herman CP. Causes of eating disorders. Annu Rev Psychol 2002; 53(1):187.
4. Levine MP, Smolak L, Hayden H. The relation of sociocultural factors to eating attitudes and behaviors among middle school girls. J Early Adolesc 1994;14(4):471–490.
5. Luff GM, Gray JJ. Complex messages regarding a thin ideal appearing in teenage girls' magazines from 1956 to 2005. Body Image 2009;6(2):133–136.
6. Field AE, Cheung L, Wolf AM, Herzog DB, Gortmaker SL, Colditz GA. Exposure to the mass media and weight concerns among girls. Pediatrics 1999;103(3):660.
7. Grabe S, Ward LM, Hyde JS. The role of the media in body image concerns among women: A meta-analysis of experimental and correlational studies. Psychol Bull 2008;134(3):460–476.
8. Groesz LM, Levine MP, Murnen SK. The effect of experimental presentation of thin media images on body satisfaction: A meta-analytic review. Int J Eat Disord 2002;31(1):1–16.
9. Field AE, Javaras KM, Aneja P, Kitos N, Camargo CA, Taylor CB, et al. Family, peer, and media predictors of becoming eating disordered. Arch Pediatr Adolesc Med 2008;162(6):574–579.
10. Neumark-Sztainer DR, Wall MM, Haines JI, Story MT, Sherwood NE, van der Berg PA. Shared risk and protective factors for overweight and disordered eating in adolescents. Am J Prev Med 2007;33(5):359–369.
11. Crandall CS. Social contagion of binge eating. J Pers Soc Psychol 1988;55(4): 588–598.
12. Zalta AK, Keel PK. Peer influence on bulimic symptoms in college students. J Abnorm Psychol 2006;115(1):185–189.
13. Keel PK, Forney KJ, Brown TA, Heatherton TF. Influence of college peers on disordered eating in women and men at 10-year follow-up. J Abnorm Psychol 2013;122(1):105–110.
14. Junco R. The relationship between frequency of Facebook use, participation in Facebook activities, and student engagement. Comput Educ 2012;58(1): 162–171.
15. iTunes. Plump&Skinny Booth for iPhone, iPod touch, and iPad on the iTunes App Store, 2012. Available at: https://itunes.apple.com/us/app/plump-slim-booth/id480524645?mt=8. Accessed June 11, 2013.
16. Tiggemann M, Slater A. NetGirls: The Internet, Facebook, and body image concern in adolescent girls. Int J Eat Disord 2013;46(6):630–633.
17. Smith AR, Hames JL, Joiner Jr. TE. Status update: Maladaptive Facebook usage predicts increases in body dissatisfaction and bulimic symptoms. J Affect Disord 2013;149(1–3):235–240.
18. Stice E. Risk and maintenance factors for eating pathology: A meta-analytic review. Psychol Bull 2002;128(5):825–848.
19. Jacobi C, Hayward C, de Zwaan M, Kraemer HC, Agras WS. Coming to terms with risk factors for eating disorders: Application of risk terminology and suggestions for a general taxonomy. Psychol Bull 2004;130(1):19–65.
20. Davis C, Katzman DK, Kaptein S, Kirsh C, Brewer H, Kalmbach K, et al. The prevalence of high-level exercise in the eating disorders: Etiological implications. Compr Psychiatry 1997;38(6):321–326.
21. Lipsey Z, Barton SB, Hulley A, Hill AJ. "After a workout. . ." Beliefs about exercise, eating and appearance in female exercisers with and without eating disorder features. Psychol Sport Exerc 2006;7(5):425–436.
22. Bulik CM, Sullivan PF, Fear JL, Joyce PR. Eating disorders and antecedent anxiety disorders: A controlled study. Acta Psychiatr Scand 1997;96(2):101–107.
23. Billingsley-Marshall RL, Basso MR, Lund BC, Hernandez ER, Johnson CL, Drevets WC, et al. Executive function in eating disorders: The role of state anxiety. Int J Eat Disord 2013;46(4):316–321.
24. Garner DM, Olmstead MP, Bohr Y, Garfinkel PE. The Eating Attitudes Test: Psychometric features and clinical correlates. Psychol Med 1982;12(04):871–878.
25. Keel PK, Baxter MG, Heatherton TF, Joiner TE. A 20-year longitudinal study of body weight, dieting, and eating disorder symptoms. J Abnorm Psychol 2007;116(2):422–432.
26. Mintz LB, O'Halloran MS. The Eating Attitudes Test: Validation with DSM-IV eating disorder criteria. J Pers Assess 2000;74(3):489.
27. Berland NW, Thompson JK, Linton PH. Correlation between the EAT-26 and the EAT-40, the Eating Disorders Inventory, and the Restrained Eating Inventory. Int J Eat Disord 1986;5(3):569–574.
28. Aitken R. Measurement of feelings using visual analogue scales. Proc R Soc Med 1969;62(10):989–993.
29. Blond A. Impacts of exposure to images of ideal bodies on male body dissatisfaction: A review. Body Image 2008;5(3):244–250.
30. Harper B, Tiggemann M. The effect of thin ideal media images on women's self-objectification, mood, and body image. Sex Roles 2008;58:649–657.
31. Haedt-Matt AA, Zalta AK, Forbush KT, Keel PK. Experimental evidence that changes in mood cause changes in body dissatisfaction among undergraduate women. Body Image 2012;9(2):216–220.
32. Speilberger CD, Gorusch RL, Lushene R, Vagg PR, Jacobs GA. The State-Trait Anxiety Inventory. Palo Alto, CA: Consulting Psychologists Press, 1983.

33. BodyIMage3D. Fat Talk Free Week. 2013. Available at: http://bi3d.tridelta.org/ourinitiatives/fattalkfreeweek. Accessed August 12, 2013.

34. McVey GL, Kirsh G, Maker D, Walker KS, Mullane J, Laliberte M, et al. Promoting positive body image among university students: A collaborative pilot study. Body Image 2010;7(3):200–204.

35. Halliwell E, Easun A, Harcourt D. Body dissatisfaction: Can a short media literacy message reduce negative media exposure effects amongst adolescent girls? Br J Health Psychol 2011;16(2):396–403.

36. Smyth J, Wonderlich S, Crosby R, Miltenberger R, Mitchell J, Rorty M. The use of ecological momentary assessment approaches in eating disorder research. Int J Eat Disord 2001;30(1):83–95.

Appendix

Facebook Questions

1. What is the average amount of time you spend on Facebook for each session? ________
2. How much overall time do you spend on Facebook each day? ________
3. How many days a week do you use Facebook? ________
4. Do you have a smart phone? Y N
   If so, do you use the Facebook application? Y N
5. If you were asked to use your Facebook in the lab, how representative was the session just now of how you normally use Facebook?
   1) Not at all
   2) Somewhat
   3) Moderately
   4) Very
   5) Completely
6. When using Facebook, which do you do the most? (Rank from 1-13, where 1 is the HIGHEST and 13 is the LOWEST.)
   1) Look at photos _____
   2) Comment on or "like" status updates _____
   3) Comment on or "like" friend's photos _____
   4) Use notes _____
   5) Use events _____
   6) Use chat or send messages _____
   7) Post your own photos _____
   8) Post your own status updates _____
   9) Find friends _____
   10) Look at business/company pages _____
   11) Use apps and games _____
   12) Use check-ins _____
   13) View or post in groups _____
7. On Facebook, what do you find to be the most interesting if you had to choose **only one**? (Please circle only one.)
   1) Look at photos
   2) Comment on or "like" status updates
   3) Comment on or "like" friend's photos
   4) Use notes
   5) Use events
   6) Use chat or send messages
   7) Post your own photos
   8) Post your own status updates
   9) Find friends
   10) Look at business/company pages
   11) Use apps and games
   12) Use check-ins
   13) View or post in groups
8. If you were asked to use your Facebook in the lab, how long ago did you use Facebook before this session? ________
9. How often do you compare your photos to photos of your female friends?
   1) Never
   2) Rarely
   3) Sometimes
   4) Usually
   5) Always
10. How important is it to you to have more likes or comments on your photos than your other female friends?
    1) Not at all
    2) Somewhat
    3) Moderately
    4) Very
    5) Extremely
11. How important is it to you that people "like" your photos?
    1) Not at all
    2) Somewhat
    3) Moderately
    4) Very
    5) Extremely
12. How important is it to you that people "like" your status updates?
    1) Not at all
    2) Somewhat
    3) Moderately
    4) Very
    5) Extremely
13. How important is it to you that people comment on your photos?
    1) Not at all
    2) Somewhat
    3) Moderately
    4) Very
    5) Extremely
14. How important is it to you that people comment on your status updates?
    1) Not at all
    2) Somewhat
    3) Moderately

4) Very
5) Extremely

15. How often do you change your profile picture?
    1) Never
    2) Once every 3 months
    3) Once a month
    4) Twice a month
    5) Once a week
    6) More than once per week
    7) Daily
16. How often do you take photos in public for the main purpose of posting them on Facebook?
    1) Never
    2) Rarely
    3) Sometimes
    4) Usually
    5) Always
17. How often do you untag your photos?
    1) Never
    2) Rarely
    3) Sometimes
    4) Usually
    5) Always
18. Why do you untag your photos?
    1) Unflattering
    2) Inappropriate for family/coworkers
    3) Not representative of who I am/what I am really like
    4) No longer dating person in photo
    5) No longer friends with person in photo
    6) Other (please specify: ______________________)

National Trends in the Prevalence and Treatment of Depression in Adolescents and Young Adults

Ramin Mojtabai, MD, PhD, MPH,[a,b] Mark Olfson, MD, MPH,[c,d] Beth Han, MD, PhD, MPH[e]

OBJECTIVES: This study examined national trends in 12-month prevalence of major depressive episodes (MDEs) in adolescents and young adults overall and in different sociodemographic groups, as well as trends in depression treatment between 2005 and 2014.

METHODS: Data were drawn from the National Surveys on Drug Use and Health for 2005 to 2014, which are annual cross-sectional surveys of the US general population. Participants included 172 495 adolescents aged 12 to 17 and 178 755 adults aged 18 to 25. Time trends in 12-month prevalence of MDEs were examined overall and in different subgroups, as were time trends in the use of treatment services.

RESULTS: The 12-month prevalence of MDEs increased from 8.7% in 2005 to 11.3% in 2014 in adolescents and from 8.8% to 9.6% in young adults (both $P < .001$). The increase was larger and statistically significant only in the age range of 12 to 20 years. The trends remained significant after adjustment for substance use disorders and sociodemographic factors. Mental health care contacts overall did not change over time; however, the use of specialty mental health providers increased in adolescents and young adults, and the use of prescription medications and inpatient hospitalizations increased in adolescents.

CONCLUSIONS: The prevalence of depression in adolescents and young adults has increased in recent years. In the context of little change in mental health treatments, trends in prevalence translate into a growing number of young people with untreated depression. The findings call for renewed efforts to expand service capacity to best meet the mental health care needs of this age group.

Departments of [a]Mental Health, Bloomberg School of Public Health, and [b]Psychiatry and Behavioral Sciences, School of Medicine, Johns Hopkins University, Baltimore, Maryland; [c]Department of Psychiatry, College of Physicians and Surgeons, Columbia University; [d]New York State Psychiatric Institute, New York, New York; and [e]Substance Abuse and Mental Health Services Administration, Rockville, Maryland

Dr Mojtabai conceptualized and designed the study, accessed data, carried the analyses, and drafted the initial manuscript; Drs Olfson and Han consulted with the research team in the design of the study and interpretation of the results, and reviewed and revised the manuscript; and all authors approved the final manuscript as submitted.

DOI: 10.1542/peds.2016-1878

Accepted for publication Aug 30, 2016

Address correspondence to Ramin Mojtabai, MD, PhD, MPH, 624 North Broadway, Rm 797, Baltimore, MD 21205. E-mail: rmojtab1@jhu.edu

PEDIATRICS (ISSN Numbers: Print, 0031-4005; Online, 1098-4275).

FINANCIAL DISCLOSURE: The authors have indicated they have no financial relationships relevant to this article to disclose.

WHAT'S KNOWN ON THIS SUBJECT: There is evidence of increased prevalence of depressive symptoms in adolescents in industrialized countries in past 3 decades. Recent suicide trends in the United States suggest that depression in adolescents and young adults may have continued to increase.

WHAT THIS STUDY ADDS: This study provides data on recent trends in major depressive episodes in adolescents and young adults overall and in sociodemographic subgroups, as well as trends in depression treatment seeking and types of treatment.

To cite: Mojtabai R, Olfson M, Han B. National Trends in the Prevalence and Treatment of Depression in Adolescents and Young Adults. *Pediatrics.* 2016;138(6):e20161878

The risk of depression sharply rises as children transition to adolescence.[1] In the US National Comorbidity Survey (NCS)-Adolescent Supplement of 2001 to 2004, 11.7% of adolescents 13 to 18 years of age met criteria for a lifetime major depressive disorder or dysthymia.[1] Reports of increasing antidepressant medication use by adolescents before the Food and Drug Administration (FDA) 2003 black-box warning[2,3] and indirect evidence of increased lifetime prevalence of major depressive disorder in successive birth cohorts[4] have raised concerns about increasing prevalence of depression among adolescents. Yet, there is little direct information from the United States on national trends in prevalence of depression in adolescents and young adults.

Studies of trends in depression from other industrialized countries have produced mixed results.[5] Although studies based on rating scales of depressive symptoms showed an increasing trend over the past 3 decades,[6–13] a 2006 meta-analysis of 26 epidemiologic studies on rates of current depressive disorder among adolescents found no significant change between the mid-1960s and mid-1990s.[14] A more recent study[15] found a declining prevalence of severe impairment among US adolescents from 1996–1998 to 2010–2012.[15] However, this study did not assess trends in specific disorders and was based on parent reports.

Examining temporal trends in prevalence of depression among young people has implications for evaluating whether they have benefited from increasing use of mental health treatments.[16] Characterization of national trends in depressive disorders and their treatment could also inform community efforts to improve access to mental health services for young people. In the current study, we used data from the 2005 to 2014 National Surveys on Drug Use and Health (NSDUH) on adolescents and young adults to examine trends in 12-month major depressive episodes (MDEs), controlling for sociodemographic characteristics and substance use disorders. We further examined trends in prevalence of MDEs in different sociodemographic groups and trends in mental health service use among adolescents and young adults with MDEs. The study period covers years following the FDA black-box warning regarding antidepressant use in youth. Stratified regression analyses are examined for trends in prevalence according to sex, race/ethnicity, age, income group, and substance use disorders. Trends in service use also are evaluated by type of provider, type of setting, use of psychiatric medications, continuation of treatment, and perceived helpfulness of treatments.

METHODS

Sample

The NSDUH is a cross-sectional annual survey of the US population in all 50 states and the District of Columbia sponsored by Substance Abuse and Mental Health Services Administration. It provides nationally representative data on MDE and its treatment among the civilian noninstitutionalized population aged 12 or older. Persons without a household address (eg, homeless persons not living in shelters), active-duty military, and institutional residents are excluded. Interviews are conducted by using computer-assisted interviewing. The NSDUH data collection protocol was approved by the institutional review board at RTI International (Research Triangle Park, NC). NSDUH oversamples adolescents and young adults.

Overall, 176 245 adolescents aged 12 to 17 and 180 459 adults aged 18 to 25 were interviewed in the NSDUH 2005 to 2014 and their data are available in public use files. The annual mean weighted response rate of the 2005 to 2014 NSDUH was 65.2%[17] according to Response Rate 2 of the American Association for Public Opinion Research.[18] Of those interviewed, 172 495 (98.9%) adolescents and 178 755 (99.1%) young adults responded to structured interview questions for 12-month MDE and comprised the study samples.

Assessments

Lifetime and 12-month MDE were assessed using a structured interview based on the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* criteria. Participants were next asked whether they had experienced an episode in the past year. Questions were adapted from the depression section of the NCS-Replication.[4] Although the validity of the NSDUH major depressive disorder instrument has not been assessed, the validity of the NCS-Replication in adults and adolescents has been assessed.[1,18] In a test-retest reliability study of NSDUH interviews, the κ values for the of past-year MDE ranged from 0.52 in adults to 0.72 in adolescents,[19] representing "moderate" to "substantial agreement," respectively.[20]

Treatments for depression were assessed by asking whether during the past 12 months participants had seen or talked to a medical doctor or other professional about their depressive symptoms. Types of professionals were aggregated into mental health providers (psychologist, psychiatrist, or psychotherapist; social worker; counselor, other mental health professional), general medical providers (general practitioner or family doctor; other medical doctor; a nurse, occupational therapist, or other health professional), and complementary/alternative medicine

(CAM) providers (eg, a religious or spiritual advisor; alternative healers, such as an herbalist). Participants were next asked if they were currently receiving treatment or counseling. In addition, participants were asked if during the past 12 months they had taken medication that was prescribed for their depression and whether they were taking such medication at the time of interview. Two further questions assessed how much treatment or counseling had helped and how much prescription medication had helped, with responses ranging from "not at all" to "extremely." Questions regarding treatment of depression were asked only among participants with positive responses to the MDE questions.

Treatment setting was assessed only in adolescents by a series of questions concerning where they received "treatment or counseling in the past 12 months for emotional or behavioral problems that were not caused by alcohol or drugs." Participants were further asked about the reason for their service use. Although these questions were not limited to participants with MDE, in our analyses we examined only settings in which adolescents had sought care because they had "felt depressed." The settings included "a private therapist, psychologist, psychiatrist, social worker, or counselor," "a mental health clinic or center," "a partial day hospital or day treatment program," and staying overnight in "any type of hospital" or in "a residential treatment center." Participants were also asked about treatment from "an in-home therapist, counselor, or family preservation worker" and from "a school social worker, a school psychologist, or a school counselor." The wording of questions regarding school services changed in 2009. Therefore, we examined this question using the 2009 to 2014 NSDUH data.

Treatment seeking for any mental health problems was assessed differently in adolescents and young adults. All adolescent participants were asked if they had received care for "emotional or behavioral problems that were not caused by alcohol or drugs." Similarly, all adult participants were asked if they had received treatment or counseling for "problems with emotions, nerves or mental health," excluding treatment of alcohol or drug use. We limited the analyses of these questions to participants with 12-month MDE.

Substance use disorder ratings were based on individual diagnostic criteria from the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* on past 12-month abuse and dependence of alcohol, marijuana, and other drugs (including cocaine, hallucinogens, heroin, inhalants, non-medically used prescription pain relievers, sedatives, tranquilizers, or stimulants). Substance use disorders were categorized into alcohol use disorder only, cannabis use disorder (without alcohol use disorder), other drug use disorder (without alcohol use disorder), and any drug use disorder with alcohol use disorder.

Information also was collected on participant age, sex, race/ethnicity, annual family income, student status, parents in the household (for adolescents), employment status (for those aged 14+) and marital status (adults).

Statistical Analyses

Analyses were conducted in 3 stages. First, trends in the prevalence of 12-month MDE across the 10 years of NSDUH were assessed by using adjusted binary logistic regression models. The adjusted model included sociodemographic variables and substance use disorders in addition to the survey year variable. Because the association between survey year and 12-month MDE did not appear to be linear in preliminary analyses, the mvrs routine of Stata 14 (Stata Corp, College Station, TX) for regression spline was used to fit the data (see Supplemental Information).[23]

Second, we conducted stratified analyses and interaction tests to examine whether the observed trends in 12-month MDE were consistent across major sociodemographic groups. The transformed predictor variables obtained in the first stage of the analyses were used as predictors in the stratified analyses.

In the third stage, we examined trends in depression treatment overall and by provider type and setting by using binary logistic models. Trends in perceived helpfulness of treatments were similarly assessed. These analyses were limited to respondents with 12-month MDE. In addition to sociodemographics and substance use disorders, these models adjusted for health insurance coverage and receipt of substance use disorder treatment.

All analyses were conducted by using Stata 14, taking into account the complex survey design and sampling weights of NSDUH; α was set at $P < .01$.

RESULTS

Of the 172 495 adolescents and 178 755 young adults who responded to 12-month MDE questions, 15 529 (8.7%) and 15 603 (8.6%), respectively, met criteria for 12-month MDE. Background characteristics by MDE are presented in Supplemental Table 4. In comparison with adolescents without MDE, those with MDE included a disproportionate number of older adolescents, nonstudents, unemployed individuals, adolescents from households with no parents or with single parents, and adolescents with substance use disorders. Adolescents with MDE also were less

likely to be boys than girls and non-Hispanic black than non-Hispanic white (Supplemental Table 4). The sociodemographic correlates of MDE in young adults were somewhat similar to those in adolescents. Compared with young adults without MDE, those with 12-month MDE were proportionately less likely to be men and non-Hispanic black compared to non-Hispanic white and more likely to have a substance use disorder. Young adults with 12-month MDE were also more likely to be unemployed or employed part-time, compared to employed full-time; widowed, divorced or separated, or never married, compared to married or living as married; and less likely to have an annual family income of \$20 000 to \$75 000, compared to <\$20,000 (Supplemental Table 4).

In adolescents, the prevalence of 12-month MDE was stable over the 2005 to 2011 period; however, it gradually increased in later years (Fig 1), growing from 8.7% (2005) to 11.3% (2014) corresponding to a 37% increase in odds (odds ratio [OR] 1.37, 95% confidence interval [CI] 1.27–1.48, $P < .001$). These proportions translate into an increase of more than a half million adolescents with 12-month MDE between 2005 (approximately 2 200 000) and 2014 (approximately 2 700 000). The change was more modest for the young adult group: from 8.8% to 9.6% (OR 1.13, 95% CI 1.05–1.22, $P = .001$).

A spline regression model with 2 basis functions provided the best-fitting model for both adolescents and young adults (see Supplemental Information). These transformed predictors were used in the regression models, the results of which are presented in Table 1.

For adolescents, a similar increasing trend was observed across age, income, and substance use disorder strata (Table 1); however, the trend was somewhat stronger among girls than boys, with the interaction

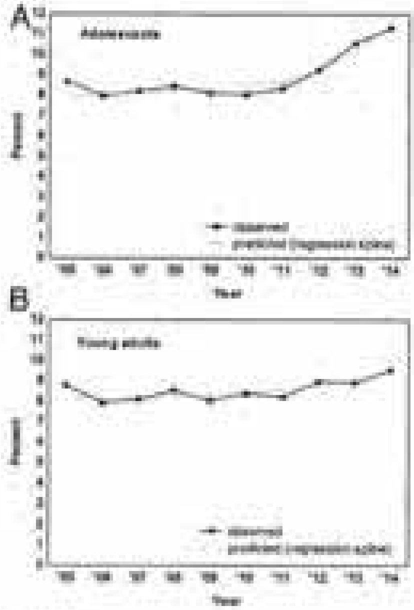


FIGURE 1
Prevalence of 12-month MDEs in adolescents (A) and young adults (B) in the United States based on the 2005 to 2014 NSDUH. The predicted value lines are based on regression spline (see text and Supplemental Information for details).

test approaching significance ($P = .020$). Among girls, the prevalence of 12-month MDE increased from 13.1% (2004) to 17.3% (2014); whereas, among boys the prevalence increased from 4.5% (2004) to 5.7% (2014) (Table 1). Furthermore, the increasing trend was smaller and statistically nonsignificant in non-Hispanic black adolescents, although the interaction test for race/ethnicity was not statistically significant (Table 1).

Among young adults, the increasing trend in prevalence of 12-month MDE was limited to those in the 18 to 20 age range (Table 1). The prevalence did not appreciably change in the 21 to 25 age range. The interaction test with age group was statistically significant ($P < .01$). Furthermore, the trend was statistically significant only among the non-Hispanic whites. However, the interaction test with race/ethnicity only approached a trend-level statistical significance ($P = .012$).

The proportion of adolescents with 12-month MDE who received mental health counseling or treatment in the past 12 months for their depression from any type of provider did not significantly change over the 2005 to 2014 period (Table 2). However, a larger proportion of adolescents with 12-month MDE reported care from specialty mental health providers, in private mental health care settings, in inpatient or day treatment settings, and in multiple settings. Furthermore, adolescents with 12-month MDE who had received any treatment or counseling in the past year were more likely in recent years to report being currently in treatment. A larger proportion of adolescents with 12-month MDE also reported receiving prescription medication for their depression in recent years. There was no significant trend in the percentage of adolescents who perceived treatment overall or medication treatment as helpful (Table 2).

There were fewer statistically significant changes in treatment seeking over time among young adults (Table 3). Only the proportion of young adults with 12-month MDE who received depression care from specialty mental health care providers increased significantly over the study period.

DISCUSSION

Each year almost 1 in 11 adolescents and young adults have an MDE. The prevalence of these episodes increased between 2005 and 2014. The trend was limited to those in the 12 to 20 age range, and was somewhat more prominent among non-Hispanic whites than minority groups and among adolescent girls than boys.

Adjusting the analyses for sociodemographic and household factors that have been previously found to be associated with adverse mental health outcomes in

TABLE 1 Trends in 12-Month MDEs in Adolescents (n = 172 495) and Young Adults (n = 178 755) in the United States, Stratified According to Major Sociodemographic Groups

	Percent of Participants in Each Survey Year Meeting Criteria for 12-mo MDEs										Main Effect[a,b]		Interaction[c]	
	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	Wald Test	P	Wald Test	P
All adolescent participants	8.7	8.0	8.2	8.4	8.0	8.0	8.3	9.3	10.6	11.3	99.82	<.001	—[d]	—[d]
Age, y														
12–13	5.6	5.3	4.3	5.1	4.5	4.2	4.2	5.6	5.9	7.2	21.15	<.001		
14–15	9.1	7.8	8.6	8.5	8.6	9.0	8.7	10.1	12.1	11.8	39.24	<.001		
16–17	11.2	10.6	11.6	11.0	10.5	10.6	11.7	11.9	13.4	14.7	39.60	<.001	1.90	.115
Sex														
Girls	13.1	11.9	12.1	12.6	11.3	12.0	12.1	14.2	16.1	17.3	98.75	<.001		
Boys	4.5	4.2	4.8	4.3	4.9	4.2	4.7	4.6	5.2	5.7	10.09	<.001	4.06	.020
Race/ethnicity														
Non-Hispanic white	9.0	8.0	8.9	8.8	8.2	8.6	8.9	9.5	10.9	12.1	60.97	<.001		
Non-Hispanic black	7.2	6.9	6.9	7.5	7.3	6.4	7.1	7.9	8.3	8.9	3.79	.026		
Hispanic	9.4	8.1	7.5	7.2	8.1	8.0	7.6	10.5	10.5	11.2	15.84	<.001		
Non-Hispanic other	7.5	9.6	6.9	10.0	7.9	6.7	8.7	7.1	12.5	10.9	7.47	<.001	0.57	.757
Family income, $														
<20 000	9.1	7.5	7.9	8.4	6.8	7.8	8.5	10.9	10.0	11.0	13.75	<.001		
20 000–49 999	9.1	9.5	9.4	8.8	8.6	8.6	9.4	9.1	10.6	11.5	13.09	<.001		
50 000–74 999	9.7	7.1	7.4	7.9	8.5	9.3	7.5	9.7	12.6	12.3	30.05	<.001		
75 000+	7.4	7.2	7.8	8.2	7.9	6.8	7.6	8.5	9.9	10.9	39.56	<.001	1.38	.228
12-mo substance use disorder														
Alcohol use disorder	21.1	17.8	19.8	23.6	21.5	22.5	23.1	26.8	31.5	29.0	12.19	<.001		
Cannabis use disorder	22.9	18.1	17.7	18.7	18.5	13.9	16.5	20.1	23.3	23.5	3.38	.038		
Other drug use disorder	24.0	26.5	35.9	34.0	35.2	38.2	28.9	43.8	35.1	39.0	3.42	.036		
Drug use disorder with alcohol use disorder	27.3	21.2	24.6	30.1	23.7	28.0	28.1	32.4	38.5	35.3	4.18	.018		
No substance use disorder	7.6	7.0	7.1	7.1	7.1	6.9	7.3	8.2	9.6	10.4	83.18	<.001	1.94	.149
All young adult participants	8.8	8.0	8.1	8.6	8.1	8.4	8.2	9.0	8.9	9.6	19.97	<.001	—[d]	—[d]
Age, y														
18–20	8.9	7.5	7.1	8.8	7.9	8.4	8.8	9.4	9.5	10.1	19.88	<.001		
21–25	8.7	8.3	8.4	8.2	8.4	7.9	7.9	8.8	8.6	9.3	4.64	.012	6.43	.002
Sex														
Women	11.9	10.3	10.8	11.6	10.6	11.8	10.9	11.9	11.9	11.8	8.07	<.001		
Men	5.7	5.7	5.5	5.5	5.5	5.2	5.6	6.1	6.0	7.4	13.03	<.001	4.01	.021
Race/ethnicity														
Non-Hispanic white	9.5	8.6	9.0	8.8	9.0	8.8	9.0	10.0	10.2	11.1	29.12	<.001		
Non-Hispanic black	6.6	7.3	6.5	6.5	6.1	8.1	5.7	5.9	5.8	6.1	0.84	.435		
Hispanic	8.1	6.2	6.5	9.2	6.7	7.6	7.4	8.5	7.4	7.8	0.22	.800		
Non-Hispanic other	8.0	7.9	7.5	8.6	7.4	7.4	9.1	8.8	9.3	9.9	3.59	.031	2.89	.012
Family income, $														
<20 000	9.3	8.4	9.2	9.1	8.8	8.9	8.6	9.7	9.6	10.0	4.34	.015		
20 000–49 999	8.8	8.0	7.9	8.5	7.7	8.3	8.4	8.8	8.6	8.9	3.41	.037		
50 000–74 999	8.5	7.0	6.4	7.5	8.1	7.0	8.1	8.0	7.5	9.1	2.90	.059		
75 000+	8.0	7.9	8.2	8.7	7.6	8.9	7.4	8.9	9.5	10.2	10.26	<.001	0.80	.570
12-mo substance use disorder														
Alcohol use disorder	14.0	15.4	14.4	14.4	14.3	13.8	13.4	17.9	17.4	17.2	5.37	.006		
Cannabis use disorder	15.0	11.3	12.7	14.9	12.8	16.0	13.2	13.3	15.4	18.3	2.47	.090		

TABLE 1 Continued

	Percent of Participants in Each Survey Year Meeting Criteria for 12-mo MDEs										Main Effect[a,b]		Interaction[c]	
	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	Wald Test	P	Wald Test	P
Other drug use disorder	24.8	19.4	16.7	24.4	19.9	21.2	23.4	18.9	22.5	28.1	1.45	.239		
Drug use disorder with alcohol use disorder	18.3	21.0	19.0	19.4	20.9	20.3	17.1	22.7	25.3	24.3	3.59	.031		
No substance use disorder	7.1	6.0	6.5	6.8	6.5	6.9	7.0	7.1	7.2	7.9	11.15	<.001	0.64	.527

Data source: The 2005–2014 NSDUH.

[a] Adjusted for age, sex, race/ethnicity, income, student status, presence of parents in household and substance use disorders in analyses for adolescents and for age, sex, race/ethnicity, income, marital status, employment, student status, and substance use disorders in analyses for young adults.

[b] Adjusted Wald F test for the predictors from the logistic regression model. In these models, 12-mo MDE was parameterized by a regression spline with 2 basis functions and 1 knot automatically selected at position 8 on the predictor variable (corresponding to year 2012) in adolescents and a regression spline with 2 basis functions and 1 knot automatically selected at position 3 (corresponding to year 2007) in young adults (see Appendix 1 for detail).

[c] Based on adjusted Wald F test for the interaction of each variable with the predictors from the spline regression model.

[d] Interaction tests examine variations in trend among subgroups and cannot be computed for whole group.

adolescents, such as single parent homes or income, did not account for the increasing trend in depression. Furthermore, the trends could not be explained by any changes in prevalence of substance use disorders as the analyses adjusted for alcohol and nonalcohol drug abuse and dependence. Analyses of NSDUH and other US surveys have not identified any meaningful increase in prevalence of substance use or misuse among adolescents over the period covered by this study.[17,22]

The trends in adolescents were different among boys and girls. This aligns with past studies that also found a larger increase in depressive symptoms in girls than boys in more recent years,[5] and recent data on trends in suicide in the United States that identified a greater increase among adolescent girls and young women.[23] Adolescent girls may have been exposed to a greater degree to depression risk factors in recent years. For example, cyberbullying may have increased more dramatically among girls than boys.[24] As compared with adolescent boys, adolescent girls also now use mobile phones with texting applications more frequently and intensively[25] and problematic mobile phone use among young people has been linked to depressed mood.[26] Interestingly, the sex differences in trends were not consistent across age groups, as the prevalence of depression followed similar temporal trends in young men and women.

The differences in trends may also represent an increase in prevalence of stressful reactions and behaviors that have similar features to depression. For example, there is some evidence that the prevalence of nonsuicidal self-injury,[27–29] which has some similar features to MDEs and may be comorbid with depressive episodes, has increased in recent years, especially in adolescent girls.[30] However, the extent to which the observed temporal trends in MDEs are confounded by such stress reactions is difficult to ascertain in NSDUH, as the survey only assessed MDEs.

The causes of the observed trends remain elusive. These trends coincided with a major economic downturn that affected the mental health of all ages in many communities.[31–35] However, consistent with a previous report of no change in the prevalence of adult MDE in the United States over a similar time period,[36] no change in the prevalence of MDE was noted in the 21 to 25 age range.

We did not observe many significant changes in mental health treatment among adolescents and young adults with 12-month MDE over the 2005 to 2014 period. The use of specialty mental health providers increased in both age groups and the use of inpatient and day treatment settings, as well as medication, increased in adolescents. Most of the increases in the use of these services were limited to the years after 2011. This is broadly consistent with previous evidence indicating that clinician diagnoses of mental disorders, use of psychotherapy, and visits to psychiatrists did not appreciably change between 2003 and 2010.[37] In view of the growing prevalence of MDE in these age groups, stable treatment rates translate into a growing number of untreated depressed adolescents. These trends suggest that little progress has been made in narrowing the mental health treatment gap for adolescent depression. This lack of progress may reflect lingering reluctance on the part of providers to diagnose and treat depression in the wake of the FDA's black-box warning regarding the use of antidepressants.[2,3]

Most adolescents receive routine primary care in pediatric settings, providing opportunities to detect and treat depression. There also

TABLE 2 Trends in Treatment Seeking in Adolescents With 12-Month MDEs in the United States (n = 15 529)

Participant Groups	Survey Year										Test of Temporal Trend		
	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	AOR	99% CI	P
Treatment seeking for any mental health problems	36.4	37.2	37.2	39.2	32.3	36.4	35.4	36.4	36.9	42.0	1.14	0.96–1.36	.043
Type of provider seen for depression care													
Specialty mental health	30.0	29.1	30.6	29.3	27.3	29.5	30.8	30.5	30.2	34.3	1.19	1.00–1.42	.009
General medical	10.2	10.4	12.3	11.3	9.7	10.5	11.5	10.3	9.8	11.7	1.04	0.77–1.40	.740
CAM	5.9	7.4	6.0	6.5	4.1	5.5	6.6	5.3	4.3	5.7	0.80	0.49–1.30	.227
Any type of provider	34.3	36.1	36.1	34.4	32.4	35.8	36.0	35.0	34.2	38.6	1.11	0.83–1.34	.127
Treatment setting for depression care													
Private mental health care provider	21.3	23.7	23.7	22.7	19.0	22.8	21.5	23.2	24.3	29.0	1.33	1.08–1.64	<.001
Mental health clinic	8.1	7.6	7.6	7.4	6.6	7.8	7.7	7.7	8.5	10.8	1.39	1.00–1.93	.010
Day treatment	3.6	4.1	4.4	3.0	3.1	3.6	5.3	5.3	5.3	5.8	1.58	1.03–2.44	.006
Inpatient	3.1	2.8	3.0	3.3	3.6	3.1	3.4	3.2	3.5	5.5	1.78	1.13–2.80	.001
Residential	2.1	2.3	1.5	1.3	1.5	1.7	2.1	2.8	1.9	3.2	1.81	0.92–3.58	.024
In home	5.5	6.7	6.3	7.4	5.4	5.7	6.0	7.5	5.6	7.9	1.25	0.85–1.82	.132
School counseling	—[a]	—[a]	—[a]	—[a]	14.9	14.3	14.8	15.8	15.6	18.6	1.63	0.92–2.89	.028
Multiple settings[b]	11.4	12.9	12.7	12.7	10.8	12.4	12.5	14.2	14.1	17.1	1.52	1.15–2.00	<.001
Use of prescription medication for depression	16.5	14.6	17.8	15.5	14.1	16.2	15.3	15.8	16.9	20.0	1.31	1.02–1.68	.006
Currently receiving treatment or counseling[c]	38.4	39.1	44.3	41.7	39.3	43.8	42.5	48.9	42.9	50.7	1.48	1.14–1.94	<.001
Currently taking prescription medication[d]	67.7	67.4	67.3	63.7	66.7	67.1	72.8	72.7	61.1	68.7	1.18	0.78–1.79	.293
Treatment or counseling helpful[e]	32.9	35.2	33.9	36.7	36.4	36.6	38.1	37.9	36.1	36.1	1.10	0.82–1.47	.367
Prescription medication helpful[f]	47.2	45.9	43.9	44.3	43.4	41.4	44.0	47.1	43.6	37.7	0.84	0.56–1.26	.261

Data source: The 2005–2014 NSDUH.

AOR, adjusted OR. ORs and AORs were estimated in binary logistic regressions with the variable of survey year transformed to a variable ranging from 0 (for 2005) and 1 (for 2014). Thus, AORs represent change in odds over the whole study period. Variables of age, sex, race/ethnicity, student status, family income, parents' presence in the household, substance use disorders, substance use treatment and any health insurance were included in the models.

[a] Questions regarding school counseling in NSDUH changed in 2009. As a result, data on school counseling for years 2005-2008 were not included in the analysis of temporal trends for this type of care.

[b] Includes any combinations of private mental health provider, mental health clinic, day hospital, inpatient, residential, and in-home settings.

[c] Among participants who received any counseling or therapy for depression in the past year.

[d] Among participants who took any prescription medication for depression in the past year.

[e] Among participants who received any counseling or therapy for depression in the past year. Helpful was defined as reporting that treatment or counseling helped "a lot" or "extremely."

[f] Among participants who took any prescription medication for depression in the past year. Helpful was defined as reporting that prescription medication helped "a lot" or "extremely."

TABLE 3 Trends in Treatment Seeking in Young Adults With 12-Month MDEs in the United States (n = 15 603)

Participant Groups	Survey Year										Test of Temporal Trend		
	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	AOR	95% CI	P
Treatment seeking for any mental health problems	29.2	24.8	25.8	35.2	29.2	27.1	28.9	29.0	29.8	29.7	1.13	0.92–1.39	.131
Type of provider seen for depression care													
Specialty mental health	30.8	29.1	37.2	37.5	29.7	31.7	31.9	32.6	33.4	34.1	1.29	1.09–1.53	.003
General medical	22.6	25.4	22.9	21.1	22.5	22.7	20.7	23.4	19.1	22.4	0.91	0.75–1.10	.266
CAM	10.2	7.8	7.8	8.7	6.9	9.0	8.1	8.8	8.6	9.6	1.06	0.73–1.53	.072
Any type of provider	44.8	43.3	41.2	40.5	43.3	45.8	44.0	46.5	45.6	48.1	1.13	0.94–1.35	.084
Use of prescription medication for depression	30.9	29.7	29.4	27.9	31.5	31.8	30.1	31.2	32.1	31.8	1.09	0.90–1.32	.343
Currently receiving treatment or counseling[a]	39.2	39.5	45.0	45.0	48.1	46.4	48.1	41.8	45.2	52.9	1.25	0.95–1.64	.036
Currently taking prescription medication[b]	55.5	61.9	67.5	66.8	72.1	67.4	69.4	64.1	65.1	73.5	1.34	0.94–1.89	.031
Treatment or counseling helpful[c]	40.4	32.2	34.6	37.1	40.2	39.0	39.6	39.8	40.4	45.2	1.19	0.91–1.54	.061
Prescription medication helpful[d]	45.8	45.4	52.5	48.1	49.3	49.9	40.6	47.2	47.9	51.1	1.15	0.83–1.61	.262

Data source: The 2005–2014 NSDUH.

AOR, adjusted OR. ORs and AORs were estimated in binary logistic regressions with the variable of survey year transformed to a variable ranging from 0 (for 2005) and 1 (for 2014). Thus, AORs represent change in odds over the whole study period. Variables of age, sex, race/ethnicity, student status, family income, marital status, employment, substance use disorders, substance use treatment and any health insurance were included in the models.

[a] Among participants who received any counseling or therapy for depression in the past year.

[b] Among participants who took any prescription medication for depression in the past year.

[c] Among participants who received any counseling or therapy for depression in the past year. Helpful was defined as reporting that treatment or counseling helped "a lot" or "extremely."

[d] Among participants who took any prescription medication for depression in the past year. Helpful was defined as reporting that prescription medication helped "a lot" or "extremely."

have been efforts by a number of professional bodies, most notably the American Academy of Pediatrics Task Force on Mental Health, to promote diagnosis and management of depression and other common mental disorders in pediatric settings and to provide resources for practice enhancements, such as diagnostic guidelines.[38–40] With the wider adoption of the medical home model under the Affordable Care Act, integrative approaches to depression treatment in pediatric primary care will likely increase.[41]

The increasing use of inpatient and day treatment settings over the 2005 to 2014 period may reflect greater intensity of treatment of adolescents. These trends may be related to enhanced mental health coverage in recent years after parity legislation, which may also explain greater continuity in treatment in more recent years.[42] However, these findings need to be corroborated in future studies, especially with regard to increased inpatient treatment that was most prominent in 2014.

The findings should be interpreted in the context of several limitations. Most notably, the study was based on self-reports by adolescents and may not correspond with parent, teacher, or clinical assessments.[15] Variations in sources of information would likely affect estimates of prevalence of disorders, but are unlikely to affect temporal trends. Furthermore, self-reports may be more sensitive than parental reports for capturing internalizing disorders, such as depression.[43,44] A second limitation is that NSDUH assessed only MDE. Temporal trends in prevalence of other common mental disorders in these age groups may be different.[5]

CONCLUSIONS

In the context of these limitations, this study provides useful

information on temporal trends in 12-month MDE and treatment of depression based on large and nationally representative samples of adolescents and young adults. Prevention, early detection, and treatment of depression and other common mental disorders in these age groups are major goals of public mental health initiatives. Yet adaptation and broad implementation of effective treatment and prevention programs remains a challenge.[45,46] The growing number of depressed adolescents and young adults who do not receive any mental health treatment for their MDE calls for renewed outreach efforts, especially in school and college health and counseling services and pediatric practices where many of the untreated adolescents and young adults with depression may be detected and managed.[47]

ABBREVIATIONS

CAM: complementary/alternative medicine
CI: confidence interval
FDA: Food and Drug Administration
MDE: major depressive episodes
NCS: National Comorbidity Survey
NSDUH: National Survey on Drug Use and Health
OR: odds ratio

FUNDING: Dr Olfson's work on this study was partly supported by National Institute on Drug Abuse R01 DA019606 and Agency for Healthcare Research and Quality U19 HS021112. No other funding was secured for this study. The findings and conclusions of this study are those of the authors and do not necessarily reflect the views of the Substance Abuse and Mental Health Services Administration or the US Department of Health and Human Services. Funded by the National Institutes of Health (NIH).

POTENTIAL CONFLICT OF INTEREST: The authors have indicated they have no potential conflicts of interest to disclose.

COMPANION PAPER: A companion paper to this article can be found online at www.pediatrics.org/cgi/doi/10.1542/peds.2016-2869.

REFERENCES

1. Merikangas KR, He JP, Burstein M, et al. Lifetime prevalence of mental disorders in U.S. adolescents: results from the National Comorbidity Survey Replication–Adolescent Supplement (NCS-A). *J Am Acad Child Adolesc Psychiatry*. 2010;49(10):980–989
2. Olfson M, Marcus SC, Druss BG. Effects of Food and Drug Administration warnings on antidepressant use in a national sample. *Arch Gen Psychiatry*. 2008;65(1):94–101
3. Zito JM, Safer DJ, DosReis S, et al. Psychotropic practice patterns for youth: a 10-year perspective. *Arch Pediatr Adolesc Med*. 2003;157(1):17–25
4. Kessler RC, Berglund P, Demler O, et al; National Comorbidity Survey Replication. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA*. 2003;289(23):3095–3105
5. Collishaw S. Annual research review: Secular trends in child and adolescent mental health. *J Child Psychol Psychiatry*. 2015;56(3):370–393
6. Collishaw S, Maughan B, Goodman R, Pickles A. Time trends in adolescent mental health. *J Child Psychol Psychiatry*. 2004;45(8):1350–1362
7. Fichter MM, Xepapadakos F, Quadflieg N, Georgopoulou E, Fthenakis WE. A comparative study of psychopathology in Greek adolescents in Germany and in Greece in 1980 and 1998–18 years apart. *Eur Arch Psychiatry Clin Neurosci*. 2004;254(1):27–35
8. Sigfusdottir ID, Asgeirsdottir BB, Sigurdsson JF, Gudjonsson GH. Trends in depressive symptoms, anxiety symptoms and visits to healthcare specialists: a national study among Icelandic adolescents. *Scand J Public Health*. 2008;36(4):361–368
9. Sweeting H, Young R, West P. GHQ increases among Scottish 15 year olds 1987–2006. *Soc Psychiatry Psychiatr Epidemiol*. 2009;44(7):579–586
10. Tick NT, van der Ende J, Verhulst FC. Ten-year trends in self-reported emotional and behavioral problems of Dutch adolescents. *Soc Psychiatry Psychiatr Epidemiol*. 2008;43(5):349–355
11. Tick NT, van der Ende J, Verhulst FC. Twenty-year trends in emotional and behavioral problems in Dutch children in a changing society. *Acta Psychiatr Scand*. 2007;116(6):473–482
12. von Soest T, Wichstrøm L. Secular trends in depressive symptoms among Norwegian adolescents from 1992 to 2010. *J Abnorm Child Psychol*. 2014;42(3):403–415
13. Fleming TM, Clark T, Denny S, et al. Stability and change in the mental health of New Zealand secondary school students 2007–2012: results from the national adolescent health surveys. *Aust N Z J Psychiatry*. 2014;48(5):472–480
14. Jane Costello E, Erkanli A, Angold A. Is there an epidemic of child or adolescent depression? *J Child Psychol Psychiatry*. 2006;47(12):1263–1271
15. Olfson M, Druss BG, Marcus SC. Trends in mental health care among children and adolescents. *N Engl J Med*. 2015;372(21):2029–2038
16. Haro JM, Arbabzadeh-Bouchez S, Brugha TS, et al. Concordance of the Composite International Diagnostic Interview Version 3.0 (CIDI 3.0) with standardized clinical assessments in the WHO World Mental Health surveys. *Int J Methods Psychiatr Res*. 2006;15(4):167–180
17. Substance Abuse and Mental Health Services Administration. National Survey on Drug Use and Health. Available at: www.samhsa.gov/data/population-data-nsduh/reports. Accessed May 20, 2016

18. American Association for Public Opinion Research. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys.* 8th ed. Lenexa, KS: American Association for Public Opinion Research; 2015
19. Substance Abuse and Mental Health Services Administration. *Reliability of Key Measures in the National Survey on Drug Use and Health.* Rockville, MD: Substance Abuse and Mental Health Services Administration, Department of Health and Human Services; 2010. HHS Publication No. SMA 09-4425
20. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med.* 2005;37(5):360–363
21. Royston P, Sauerbrei W. *Multivariable Model-Building: A Pragmatic Approach to Regression Analysis Based on Fractional Polynomials for Modelling Continuous Variables.* Chichester, UK: John Wiley & Sons; 2008
22. Johnston LD, O'Malley PM, Miech RA, Bachman JG, Schulenberg JE. *Monitoring the Future National Survey Results on Drug Use, 1975–2015: Overview, Key Findings on Adolescent Drug Use.* Ann Arbor, MI: Institute for Social Research, The University of Michigan; 2016
23. Curtin SC, Warner M, Hedegaard H. *Increase in Suicide in the United States, 1999–2014.* Hyattsville, MD: National Center for Health Statistics; 2016
24. Kessel Schneider S, O'Donnell L, Smith E. Trends in cyberbullying and school bullying victimization in a regional census of high school students, 2006–2012. *J Sch Health.* 2015;85(9):611–620
25. Lenhart A. *Teen, Social Media and Technology Overview.* Washington, DC: Pew Research Center; 2015
26. Augner C, Hacker GW. Associations between problematic mobile phone use and psychological parameters in young adults. *Int J Public Health.* 2012;57(2):437–441
27. Fox KR, Franklin JC, Ribeiro JD, Kleiman EM, Bentley KH, Nock MK. Meta-analysis of risk factors for nonsuicidal self-injury. *Clin Psychol Rev.* 2015;42:156–167
28. Whitlock J, Muehlenkamp J, Purington A, et al. Nonsuicidal self-injury in a college population: general trends and sex differences. *J Am Coll Health.* 2011;59(8):691–698
29. Nock M. *The Oxford Handbook of Suicide and Self-injury.* New York, NY: Oxford University Press; 2014
30. Muehlenkamp JJ, Williams KL, Gutierrez PM, Claes L. Rates of non-suicidal self-injury in high school students across five years. *Arch Suicide Res.* 2009;13(4):317–329
31. Barr B, Kinderman P, Whitehead M. Trends in mental health inequalities in England during a period of recession, austerity and welfare reform 2004 to 2013. *Soc Sci Med.* 2015;147:324–331
32. Katikireddi SV, Niedzwiedz CL, Popham F. Trends in population mental health before and after the 2008 recession: a repeat cross-sectional analysis of the 1991–2010 Health Surveys of England. *BMJ Open.* 2012;2(5):e001790
33. Solantaus T, Leinonen J, Punamaki RL. Children's mental health in times of economic recession: replication and extension of the family economic stress model in Finland. *Dev Psychol.* 2004;40(3):412–429
34. National Academies of Sciences, Engineering, and Medicine. *Mental Disorders and Disabilities Among Low-Income Children.* Washington, DC: The National Academies Press; 2015
35. Bachmann CJ, Aagaard L, Burcu M, et al. Trends and patterns of antidepressant use in children and adolescents from five western countries, 2005–2012. *Eur Neuropsychopharmacol.* 2016;26(3):411–419
36. Mojtabai R, Jorm AF. Trends in psychological distress, depressive episodes and mental health treatment-seeking in the United States: 2001–2012. *J Affect Disord.* 2015;174:556–561
37. Olfson M, Blanco C, Wang S, Laje G, Correll CU. National trends in the mental health care of children, adolescents, and adults by office-based physicians. *JAMA Psychiatry.* 2014;71(1):81–90
38. Foy JM; American Academy of Pediatrics Task Force on Mental Health. Enhancing pediatric mental health care: report from the American Academy of Pediatrics Task Force on Mental Health. Introduction. *Pediatrics.* 2010;125(suppl 3):S69–S74
39. Foy JM; American Academy of Pediatrics Task Force on Mental Health. Enhancing pediatric mental health care: algorithms for primary care. *Pediatrics.* 2010;125(suppl 3):S109–S125
40. Foy JM, Kelleher KJ, Laraque D; American Academy of Pediatrics Task Force on Mental Health. Enhancing pediatric mental health care: strategies for preparing a primary care practice. *Pediatrics.* 2010;125(suppl 3):S87–S108
41. Ader J, Stille CJ, Keller D, Miller BF, Barr MS, Perrin JM. The medical home and integrated behavioral health: advancing the policy agenda. *Pediatrics.* 2015;135(5):909–917
42. Grazier KL, Eisenberg D, Jedele JM, Smiley ML. Effects of mental health parity on high utilizers of services: pre-post evidence from a large, self-insured employer. *Psychiatr Serv.* 2016;67(4):448–451
43. Sourander A, Helstelä L, Helenius H. Parent-adolescent agreement on emotional and behavioral problems. *Soc Psychiatry Psychiatr Epidemiol.* 1999;34(12):657–663
44. Martin JL, Ford CB, Dyer-Friedman J, Tang J, Huffman LC. Patterns of agreement between parent and child ratings of emotional and behavioral problems in an outpatient clinical setting: when children endorse more problems. *J Dev Behav Pediatr.* 2004;25(3):150–155
45. Aarons GA, Palinkas LA. Implementation of evidence-based practice in child welfare: service provider perspectives. *Adm Policy Ment Health.* 2007;34(4):411–419
46. Hoagwood K, Burns BJ, Kiser L, Ringeisen H, Schoenwald SK. Evidence-based practice in child and adolescent mental health services. *Psychiatr Serv.* 2001;52(9):1179–1189
47. McCabe MA. *Health Care Reform as a Vehicle for Promoting Children's Mental and Behavioral Health.* Washington, DC: Institute of Medicine; 2015

JAMA Psychiatry | Original Investigation

Associations Between Time Spent Using Social Media and Internalizing and Externalizing Problems Among US Youth

Kira E. Riehm, MS; Kenneth A. Feder, PhD; Kayla N. Tormohlen, MPH; Rosa M. Crum, MD; Andrea S. Young, PhD; Kerry M. Green, PhD; Lauren R. Pacek, PhD; Lareina N. La Flair, PhD; Ramin Mojtabai, MD

Author Audio Interview

Supplemental content

IMPORTANCE Social media use may be a risk factor for mental health problems in adolescents. However, few longitudinal studies have investigated this association, and none have quantified the proportion of mental health problems among adolescents attributable to social media use.

OBJECTIVE To assess whether time spent using social media per day is prospectively associated with internalizing and externalizing problems among adolescents.

DESIGN, SETTING, AND PARTICIPANTS This longitudinal cohort study of 6595 participants from waves 1 (September 12, 2013, to December 14, 2014), 2 (October 23, 2014, to October 30, 2015), and 3 (October 18, 2015, to October 23, 2016) of the Population Assessment of Tobacco and Health study, a nationally representative cohort study of US adolescents, assessed US adolescents via household interviews using audio computer-assisted self-interviewing. Data analysis was performed from January 14, 2019, to May 22, 2019.

EXPOSURES Self-reported time spent on social media during a typical day (none, ≤30 minutes, >30 minutes to ≤3 hours, >3 hours to ≤6 hours, and >6 hours) during wave 2.

MAIN OUTCOM
Harms
alone,
externalizing
lems during
wave 3 using

RESULTS A to
3%] male)
were studied.
n social
media, compa
g problems
alone (≤30 m
tes to ≤3
hours: RRR, 1.
9; >6 hours:
RRR, 2.83; 95
blems
(≤30 minutes
4; 95% CI,
1.83-3.00; >3
5% CI,
3.22-5.73); as
ted analyses,
use of social
ed
significantly a
, 1.60; 95% CI,
1.11-2.31; >6 h
xternalizing
problems (>3
% CI,
1.73-3.43) but

CONCLUSIONS AND RELEVANCE Adolescents who spend more than 3 hours per day using social media may be at heightened risk for mental health problems, particularly internalizing problems. Future research should determine whether setting limits on daily social media use, increasing media literacy, and redesigning social media platforms are effective means of reducing the burden of mental health problems in this population.

Author Affiliations: Author affiliations are listed at the end of this article.

Corresponding Author: Kira E. Riehm, MS, Department of Mental Health, Bloomberg School of Public Health, Johns Hopkins University, 624 N Broadway, Baltimore, MD 21205 (kriehm@jhu.edu).

JAMA Psychiatry. doi:10.1001/jamapsychiatry.2019.2325
Published online September 11, 2019.

JAMA Psychiatry | Original Investigation

Associations Between Time Spent Using Social Media and Internalizing and Externalizing Problems Among US Youth

Kira E. Riehm, MS; Kenneth A. Feder, PhD; Kayla N. Tormohlen, MPH; Rosa M. Crum, MD; Andrea S. Young, PhD; Kerry M. Green, PhD; Lauren R. Pacek, PhD; Lareina N. La Flair, PhD; Ramin Mojtabai, MD

Author Audio Interview

Supplemental content

IMPORTANCE Social media use may be a risk factor for mental health problems in adolescents. However, few longitudinal studies have investigated this association, and none have quantified the proportion of mental health problems among adolescents attributable to social media use.

OBJECTIVE To assess whether time spent using social media per day is prospectively associated with internalizing and externalizing problems among adolescents.

DESIGN, SETTING, AND PARTICIPANTS This longitudinal cohort study of 6595 participants from waves 1 (September 12, 2013, to December 14, 2014), 2 (October 23, 2014, to October 30, 2015), and 3 (October 18, 2015, to October 23, 2016) of the Population Assessment of Tobacco and Health study, a nationally representative cohort study of US adolescents, assessed US adolescents via household interviews using audio computer-assisted self-interviewing. Data analysis was performed from January 14, 2019, to May 22, 2019.

EXPOSURES Self-reported time spent on social media during a typical day (none, ≤30 minutes, >30 minutes to ≤3 hours, >3 hours to ≤6 hours, and >6 hours) during wave 2.

MAIN OUTCOMES AND MEASURE Self-reported past-year internalizing problems alone, externalizing problems alone, and comorbid internalizing and externalizing problems during wave 3 using the Global Appraisal of Individual Needs-Short Screener.

RESULTS A total of 6595 adolescents (aged 12-15 years during wave 1; 3400 [51.3%] male) were studied. In unadjusted analyses, spending more than 30 minutes of time on social media, compared with no use, was associated with increased risk of internalizing problems alone (≤30 minutes: relative risk ratio [RRR], 1.30; 95% CI, 0.94-1.78; >30 minutes to ≤3 hours: RRR, 1.89; 95% CI, 1.36-2.64; >3 to ≤6 hours: RRR, 2.47; 95% CI, 1.74-3.49; >6 hours: RRR, 2.83; 95% CI, 1.88-4.26) and comorbid internalizing and externalizing problems (≤30 minutes: RRR, 1.39; 95% CI, 1.06-1.82; >30 minutes to ≤3 hours: RRR, 2.34; 95% CI, 1.83-3.00; >3 to ≤6 hours: RRR, 3.15; 95% CI, 2.43-4.09; >6 hours: RRR, 4.29; 95% CI, 3.22-5.73); associations with externalizing problems were inconsistent. In adjusted analyses, use of social media for more than 3 hours per day compared with no use remained significantly associated with internalizing problems alone (>3 to ≤6 hours: RRR, 1.60; 95% CI, 1.11-2.31; >6 hours: RRR, 1.78; 95% CI, 1.15-2.77) and comorbid internalizing and externalizing problems (>3 to ≤6 hours: RRR, 2.01; 95% CI, 1.51-2.66; >6 hours: RRR, 2.44; 95% CI, 1.73-3.43) but not externalizing problems alone.

CONCLUSIONS AND RELEVANCE Adolescents who spend more than 3 hours per day using social media may be at heightened risk for mental health problems, particularly internalizing problems. Future research should determine whether setting limits on daily social media use, increasing media literacy, and redesigning social media platforms are effective means of reducing the burden of mental health problems in this population.

JAMA Psychiatry. doi:10.1001/jamapsychiatry.2019.2325
Published online September 11, 2019.

Author Affiliations: Author affiliations are listed at the end of this article.

Corresponding Author: Kira E. Riehm, MS, Department of Mental Health, Bloomberg School of Public Health, Johns Hopkins University, 624 N Broadway, Baltimore, MD 21205 (kriehm@jhu.edu).

For adolescents in the United States, social media use is ubiquitous. A 2018 Pew Research Center poll found that 97% of adolescents report using at least 1 of the 7 most popular social media platforms (YouTube, Instagram, Snapchat, Facebook, Twitter, Tumblr, and Reddit). Moreover, digital media use by adolescents is common: 95% report owning or having access to a smartphone, and almost 90% report they are online at least several times a day.[1]

Social media offers numerous potential benefits to users, including exposure to current events, interpersonal connection, and enhancement of social support networks.[2] However, concerns are increasingly raised about potential harms of social media use.[3] One-quarter of adolescents think social media has a mostly negative influence on people their age, pointing to reasons like rumor spreading, lack of in-person contact, unrealistic views of others' lives, peer pressure, and mental health issues.[1]

An increasing body of literature suggests that social media use is associated with mental health problems in adolescence. Numerous cross-sectional studies and a limited number of longitudinal studies suggest that high levels of social media use are associated with internalizing problems, including depressive and anxiety symptoms,[3-6] although results are not entirely consistent.[7] Some studies also suggest an association between social media use and externalizing problems, such as bullying and attention problems.[8,9] Furthermore, a previous study[4] produced mixed results regarding the possible moderating effect of sex.

The prevalence of major depressive disorder and depressive symptoms has increased among adolescents in the United States,[10,11] and adolescent suicide death and attempt rates have increased sharply during the past 2 decades.[12,13] Some authors[14] have postulated that increases in depression may be attributable to rapid increases in social media use. However, evidence of this association in nationally representative samples is scarce, and little is known about whether reducing time spent on social media might influence the prevalence of mental health problems at a national level.

In this article, we build on existing literature by examining the prospective association of time spent on social media with internalizing and externalizing problems in a representative sample of US adolescents. We used data from the Population Assessment of Tobacco and Health (PATH) study, which is a nationally representative, longitudinal cohort of adolescents.[15] Unlike a prior study,[16] we adjusted for mental health problems measured before the exposure, which is critical for reducing the influence of reverse causality. We hypothesized that greater time spent on social media would prospectively be associated with internalizing and externalizing problems alone, as well as comorbid problems at 1-year follow-up. On the basis of past research,[3] we also examined whether these associations differed between males and females.

Key Points

Question Is time spent using social media associated with mental health problems among adolescents?

Findings In this cohort study of 6595 US adolescents, increased time spent using social media per day was prospectively associated with increased odds of reporting high levels of internalizing and comorbid internalizing and externalizing problems, even after adjusting for history of mental health problems.

Meaning Adolescents who spend more than 3 hours per day on social media may be at heightened risk for mental health problems, particularly internalizing problems.

Methods

Participants

In this longitudinal cohort study, participants were drawn from the public-use data files of waves 1 (September 12, 2013, to December 14, 2014), 2 (October 23, 2014, to October 30, 2015), and 3 (October 18, 2015, to October 23, 2016) of the PATH study.[15] The methods of the PATH study have been previously described.[15] In brief, the target population for this survey was the civilian household population in the United States. Data were collected in 1-year intervals, starting with wave 1 from September 12, 2013, to December 14, 2014. Multistage-stratified sampling was used to obtain a sample of households from which up to 2 individuals aged 12 to 17 years were randomly selected to be interviewed. Data analysis was performed from January 14, 2019, to May 22, 2019. After oral parent permission and adolescent assent were obtained, adolescents were interviewed using audio computer-assisted self-interviewing. The current analyses were considered exempt from human subjects research according to Johns Hopkins institutional review board policy because the data were publicly available and deidentified.

The weighted response rate for adolescents during wave 1 was 78.4%, and the weighted retention rate during wave 3 was 83.3%.[17] A total of 7595 adolescents (aged 12-15 years during wave 1, aged 13-16 years during wave 2, and aged 14-17 years during wave 3) completed all 3 PATH survey waves. Of these, 1000 adolescents (13.2%) were excluded because they were missing data on at least 1 variable required for this analysis; the remaining 6595 adolescents comprised the analytic sample (eFigure in the Supplement).

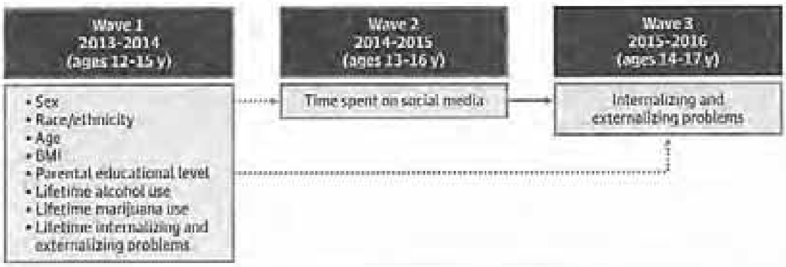
Measures

Outcome (Wave 3)

Past-year mental health problems, the outcome of interest, were assessed during wave 3 using the Global Appraisal of Individual Needs-Short Screener (GAIN-SS).[18] The GAIN-SS is a screening measure intended to identify a probable mental health disorder and assess symptom severity; it has been validated in adolescents[19] and includes internalizing and externalizing subscales (eTable 1 in the Supplement). Each item measures 1 symptom; for this study, symptoms were considered to be present if the respondent selected in the past month or 2 to 12 months from the response options that indicated the last time they had experienced that symptom. Symptom counts were generated for each subscale. Adolescents were classified as reporting low to moderate (0-3 symptoms) or high (≥4 symptoms) internalizing and externalizing problems. These cut points have been validated for use when making treatment

[redacted]

Figure 1. Directed Acyclic Graph of the Hypothesized Associations Between Study Variables and Waves of Measurement for the Exposure, Outcome, and Potential Confounders



The dashed lines represent potential confounding. The solid line represents the main association of interest. BMI indicates body mass index.

decisions[18] and have previously been used with the PATH sample.[20,21] We combined these subscales to create a single outcome variable with 4 mutually exclusive categories: no or low internalizing and externalizing problems, internalizing problems alone, externalizing problems alone, and comorbid internalizing and externalizing problems. Comorbid problems were defined as having all 4 internalizing and 4 or more externalizing symptoms.

Exposure (Wave 2)

The exposure of interest was time spent using social media per day during wave 2. Adolescents who reported that they ever went online were asked, "Sometimes people use the internet to connect with other people online through social networks like Facebook, Google Plus, YouTube, MySpace, Linkedin, Twitter, Tumblr, Instagram, Pinterest, or Snapchat. This is often called 'social media.' Do you have a social media account?" Adolescents who reported that they had a social media account that they visited were asked, "On a typical day, about how much total time do you spend on social media sites?" The response options were up to 30 minutes; more than 30 minutes, up to 3 hours; more than 3 hours, up to 6 hours; and more than 6 hours. We retained these categories for our exposure variable, with an additional category of none for adolescents who reported not going online, not having a social media account, or never visiting their social media account.

Covariates (Wave 1)

Potential confounders, including demographic characteristics (ie, sex, age, race, and parental educational level), body mass index (based on parent-reported weight and height), self-reported lifetime marijuana use and alcohol use, and scale scores for lifetime internalizing and externalizing problems, were adjusted for in the analyses. To ensure that we did not improperly adjust for mediating variables,[22] we used covariates measured at wave 1 instead of wave 2. The full study design is displayed in **Figure 1**.

Statistical Analysis

Multinomial logistic regression was used to estimate the associations between time spent on social media per day with internalizing problems alone, externalizing problems alone, and comorbid internalizing and externalizing problems (reference group: no or low internalizing and externalizing problems). Both unadjusted and adjusted analyses were conducted. Regression coefficients were exponentiated for interpretation as relative risk ratios (RRRs). In addition, we used the adjusted model to generate and plot predicted probabilities of high internalizing and externalizing problems for each level of social media use for an otherwise average study participant.

We tested for the presence of a linear trend in the coefficients for social media use in their relation to each category of mental health problems by converting the social media use variable to an ordinal variable and reestimating the adjusted model (ie, a Mantel test for trend[23]). A linear trend would suggest that more time spent on social media is associated with a proportionally greater likelihood of reporting mental health problems.

We tested whether any observed association of social media use with mental health problems differed between males and females by testing an interaction term between social media use and sex in our adjusted model.

In addition, we estimated the respective proportions of high internalizing and high externalizing problem cases that would be potentially prevented if adolescents spent less time using social media (ie, the population-attributable fraction [PAF] for social media use). We did this for 4 counterfactual scenarios that represented increasingly greater population reductions in social media use. In scenario 1, adolescents who actually used social media more than 6 hours per day would instead use social media more than 3 hours to 6 hours or less per day; in scenario 2, adolescents who actually used social media more than 3 hours per day would instead use social media more than 30 minutes to 3 hours or less per day; in scenario 3, adolescents who actually used social media more than 30 minutes per day would instead use social media 30 minutes or less per day; and in scenario 4, adolescents who actually spent any amount of time on social media per day would instead not spend any time on social media.

We estimated each scenario by generating a counterfactual population from our adjusted model using the approach to calculate PAFs described by Greenland and Drescher[24] and Rückinger et al.[25] See the eMethods in the Supplement for a detailed description.

To test whether our results were sensitive to missing data, we repeated analyses using multiply imputed data. We performed multiple imputation using chained equations and recomputed the unadjusted, adjusted, and sex-interaction models. We stratified by sex and generated 10 imputed data sets

Table 1. Descriptive Statistics of Population Characteristics for US Adolescents in the PATH Study, 2013-2016, Overall and by Internalizing and Externalizing Problems[a]

Variable	Total Sample (N = 6595)	Internalizing Problems Alone During Wave 3 (n = 611)	Externalizing Problems Alone During Wave 3 (n = 885)	Internalizing and Externalizing Problems During Wave 3 (n = 1169)
Time spent on social media per day during wave 2				
None	1125 (16.8)	73 (6.6)	122 (12.4)	122 (10.7)
≤30 min	2082 (31.8)	172 (7.8)	287 (14.4)	283 (11.6)
>30 min to ≤3 h	2000 (30.7)	198 (9.6)	310 (15.5)	390 (19.6)
>3 to ≤6 h	817 (12.3)	98 (11.9)	97 (12.2)	202 (24.6)
>6 h	571 (8.4)	70 (12.1)	69 (12.7)	172 (29.7)
Sex				
Male	3400 (51.3)	180 (5.0)	564 (17.3)	423 (12.5)
Female	3195 (48.7)	431 (13.4)	321 (10.5)	746 (23.1)
Race				
White only	4563 (70.9)	432 (9.3)	635 (14.5)	831 (18.3)
Black only	1000 (14.8)	68 (6.9)	131 (13.4)	147 (15.0)
Other[b]	1032 (14.3)	111 (10.2)	119 (12.3)	191 (17.2)
Parental educational level				
Less than high school	1308 (17.0)	116 (8.7)	137 (11.0)	183 (14.4)
High school or equivalent	1218 (17.8)	140 (11.3)	133 (11.0)	217 (18.1)
Some college or associate's degree	2072 (31.0)	202 (9.4)	290 (14.3)	417 (19.9)
Bachelor's degree	1296 (21.8)	103 (8.1)	196 (15.7)	232 (17.1)
Advanced degree	701 (12.5)	50 (6.8)	126 (18.7)	120 (16.9)
Age, y				
12-14	4913 (74.2)	443 (8.9)	662 (14.0)	888 (17.8)
15-17	1682 (25.8)	168 (9.5)	223 (13.9)	281 (17.4)
BMI, mean (SD)	21.91 (5.03)	22.30 (5.25)	21.53 (4.57)	22.18 (5.00)
Lifetime alcohol use				
No	4661 (70.0)	410 (8.6)	590 (13.4)	688 (14.5)
Yes	1934 (30.0)	201 (10.2)	295 (15.5)	481 (25.1)
Lifetime marijuana use				
No	6132 (93.3)	561 (8.9)	826 (14.1)	1062 (17.3)
Yes	463 (6.7)	50 (11.2)	59 (12.8)	107 (23.1)
No. of lifetime internalizing problems, mean (SD)	2.19 (1.57)	2.84 (1.37)	2.38 (1.44)	3.19 (1.23)
No. of lifetime externalizing problems, mean (SD)	3.22 (2.12)	3.33 (1.94)	4.08 (1.83)	4.49 (1.78)

Abbreviations: BMI, body mass index (calculated as weight in kilograms divided by height in meters squared); PATH, Population Assessment of Tobacco and Health.

[a] Data are presented as number (percentage) of patients unless otherwise indicated. Percentages, means, and SDs are weighted using the wave 3 all-waves replicate weights. All variables were measured during wave 1 except time spent on social media per day, which was measured during wave 2.

[b] The other race category includes participants identifying as American Indian or Alaska Native, Asian Indian, Chinese, Filipino, Japanese, Korean, Vietnamese, other Asian, Native Hawaiian, Guamanian or Chamorro, Samoan, and other Pacific Islander.

to account for the hypothesized interaction between sex and social media use.[39]

Data for analyses were weighted to be representative of 12- to 15-year-old adolescents living in the United States in 2013 to 2014. Standard errors were estimated using the wave 3 all-waves replicate weights constructed using balanced repeated replication (the Fay method) provided in the PATH data set. Statistical significance was assessed at a 2-sided $P < .05$ level. All analyses were conducted using Stata, version 14 (StataCorp).

Results

Sample Characteristics

A total of 6595 adolescents (aged 12-15 years during wave 1; 3400 [51.3%] male) were included in the analysis. During wave 3, of the sample of 6595 adolescents, 611 (9.1%) reported internalizing problems alone, 885 (14.0%) reported externalizing problems alone, 1169 (17.7%) reported comorbid internalizing and externalizing problems, and the remaining 3930 (59.3%) reported no or low problems. During wave 2, a total of 1125 adolescents (16.8%) reported no social media use, 2082 (31.8%) reported 30 minutes or less, 2000 (30.7%) reported more than 30 minutes to 3 hours or more, 817 (12.3%) reported more than 3 hours to 6 hours or less, and 571 (8.4%) reported more than 6 hours of use per day. Sample characteristics are given in **Table 1**.

Association Between Social Media Use and Mental Health Problems

Compared with adolescents who did not use social media, the use of social media for more than 30 minutes per day was as-

Table 2. Unadjusted and Adjusted RRRs for Each Category of Social Media Use Associated With Internalizing and Externalizing Problems Among 6595 US Adolescents in the PATH Study, 2013-2016[a]

Variable	Internalizing Problems Alone		Externalizing Problems Alone		Comorbid Internalizing and Externalizing Problems	
	RRR (95% CI)	aRRR (95% CI)	RRR (95% CI)	aRRR (95% CI)	RRR (95% CI)	aRRR (95% CI)
Time spent on social media per day during wave 2						
None	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]
≤30 min	1.30 (0.94-1.78)	1.23 (0.89-1.71)	1.28 (0.98-1.67)	1.18 (0.89-1.56)	1.39 (1.06-1.82)	1.27 (0.97-1.67)
>30 min to ≤3 h	1.89 (1.36-2.64)	1.37 (0.96-1.94)	1.60 (1.16-2.21)	1.37 (0.98-1.92)	2.34 (1.83-3.00)	1.59 (1.23-2.05)
>3 to ≤6 h	2.47 (1.74-3.49)	1.60 (1.11-2.31)	1.36 (0.97-1.90)	1.22 (0.86-1.72)	3.15 (2.43-4.09)	2.01 (1.51-2.66)
>6 h	2.83 (1.88-4.26)	1.78 (1.15-2.77)	1.59 (1.07-2.37)	1.40 (0.90-2.19)	4.29 (3.22-5.73)	2.44 (1.73-3.43)
Sex						
Male	NA	0.38 (0.30-0.47)	NA	1.25 (1.03-1.53)	NA	0.51 (0.43-0.61)
Female	NA	1 [Reference]	NA	1 [Reference]	NA	1 [Reference]
Race						
White only	NA	1 [Reference]	NA	1 [Reference]	NA	1 [Reference]
Black only	NA	0.65 (0.50-0.83)	NA	0.86 (0.67-1.10)	NA	0.70 (0.54-0.91)
Other[b]	NA	1.00 (0.73-1.36)	NA	0.85 (0.67-1.09)	NA	0.86 (0.68-1.09)
Parental educational level						
Less than high school	NA	1 [Reference]	NA	1 [Reference]	NA	1 [Reference]
High school or equivalent	NA	1.38 (1.05-1.82)	NA	0.99 (0.75-1.31)	NA	1.23 (0.93-1.63)
Some college or associate's degree	NA	1.17 (0.90-1.51)	NA	1.29 (1.02-1.63)	NA	1.37 (1.08-1.75)
Bachelor's degree	NA	0.99 (0.72-1.34)	NA	1.34 (0.99-1.81)	NA	1.18 (0.89-1.57)
Advanced degree	NA	0.89 (0.60-1.32)	NA	1.69 (1.24-2.31)	NA	1.28 (0.91-1.79)
Age, y						
12-14	NA	1 [Reference]	NA	1 [Reference]	NA	1 [Reference]
15-17	NA	0.94 (0.77-1.14)	NA	0.94 (0.79-1.12)	NA	0.82 (0.70-0.96)
BMI	NA	1.00 (0.98-1.02)	NA	0.99 (0.97-1.00)	NA	1.00 (0.98-1.01)
Lifetime alcohol use						
No	NA	1 [Reference]	NA	1 [Reference]	NA	1 [Reference]
Yes	NA	1.02 (0.84-1.25)	NA	0.97 (0.82-1.14)	NA	1.17 (1.00-1.36)
Lifetime marijuana use						
No	NA	1 [Reference]	NA	1 [Reference]	NA	[1 [Reference]
Yes	NA	0.94 (0.65-1.37)	NA	0.67 (0.47-0.95)	NA	0.71 (0.54-0.95)
Lifetime internalizing problems	NA	1.57 (1.45-1.71)	NA	1.00 (0.93-1.07)	NA	1.48 (1.38-1.60)
Lifetime externalizing problems	NA	0.97 (0.91-1.03)	NA	1.43 (1.35-1.51)	NA	1.36 (1.27-1.44)

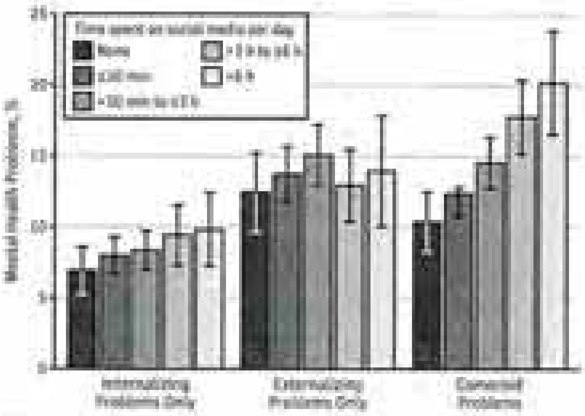
Abbreviations: aRRR, adjusted relative risk ratio; BMI, body mass index (calculated as weight in kilograms divided by height in meters squared); NA, not applicable; PATH, Population Assessment of Tobacco and Health; RRR, relative risk ratio.

[a] The aRRRs are adjusted for all covariates listed in Table 1. The reference category is no internalizing or externalizing problems. All variables were measured during wave 1 except time spent on social media per day, which was measured during wave 2.

[b] The other race category includes participants identifying as American Indian or Alaska Native, Asian Indian, Chinese, Filipino, Japanese, Korean, Vietnamese, other Asian, Native Hawaiian, Guamanian or Chamorro, Samoan, and other Pacific Islander.

sociated with greater risk of internalizing problems alone (≤30 minutes: RRR, 1.30; 95% CI, 0.94-1.78; >30 minutes to ≤3 hours: RRR, 1.89; 95% CI, 1.36-2.64; >3 to ≤6 hours: RRR, 2.47; 95% CI, 1.74-3.49; >6 hours: RRR, 2.83; 95% CI, 1.88-4.26) and comorbid internalizing and externalizing problems (≤30 minutes: RRR, 1.39; 95% CI, 1.06-1.82; >30 minutes to ≤3 hours: RRR, 2.34; 95% CI, 1.83-3.00; >3 to ≤6 hours: RRR, 3.15; 95% CI, 2.43-4.09; >6 hours: RRR, 4.29; 95% CI, 3.22-5.73) (Table 2). In the adjusted model, the associations for the 2 highest categories of social media use persisted for internalizing problems alone (>3 to ≤6 hours: RRR, 1.60; 95% CI, 1.11-2.31; >6 hours: RRR, 1.78; 95% CI, 1.15-2.77), and the associations for the 3 highest categories of social media use persisted for comorbid internalizing and externalizing problems (>30 minutes to ≤3 hours: RRR, 1.59; 95% CI, 1.23-2.05; >3 to ≤6 hours: RRR, 2.01; 95% CI, 1.51-2.66; >6 hours: RRR, 2.44; 95% CI, 1.73-3.43). In contrast, in unadjusted analyses, the association of social media use with externalizing problems was inconsistent (≤30 minutes: RRR, 1.28; 95% CI, 0.98-1.67; >30 minutes to ≤3 hours: RRR, 1.60; 95% CI, 1.16-2.21; >3 to ≤6 hours: RRR,

Figure 2. Adjusted Proportion of Internalizing Problems, Externalizing Problems, and Comorbid Internalizing and Externalizing Problems Stratified by Category of Time Spent on Social Media per Day Among US Adolescents in the Population Assessment of Tobacco and Health Study, 2013-2016



Error bars indicate 95% CIs.

1.36; 95% CI, 0.97-1.90; >6 hours: RRR, 1.59; 95% CI, 1.07-2.37) and not significant in the adjusted analysis (≤30 minutes: RRR, 1.18; 95% CI, 0.89-1.56; >30 minutes to ≤3 hours: RRR, 1.37; 95% CI, 0.98-1.92; >3 to ≤6 hours: RRR, 1.22; 95% CI, 0.86-1.72; >6 hours: RRR, 1.40; 95% CI, 0.90-2.19) (Table 2). The predicted probabilities of high internalizing, externalizing, and comorbid problems for each level of social media use, with all other covariates set to their mean, are displayed in **Figure 2**.

We observed a significant linear trend in the coefficients for both internalizing ($F_{1,96}$ = 8.86, P = .004) and comorbid problems ($F_{1,96}$ = 35.16, P < .001); as time on social media increased, the odds of these outcomes increased proportionately. In contrast, we observed no association for externalizing problems ($F_{1,96}$ = 2.25, P = .14).

We observed no statistically significant interaction between social media use and sex for internalizing ($F_{4,96}$ = 0.84, P = .50), externalizing ($F_{4,96}$ = 0.32, P = .86), or comorbid problems ($F_{4,96}$ = 0.73, P = .57).

All PAF estimates are given in **Table 3**. On the basis of our adjusted model assuming no confounding, 0.8% to 18.9% of internalizing problems and 0.8% to 15.3% of externalizing problems could be prevented if participants had instead used less social media.

Results of analyses using multiple imputation methods did not differ appreciably from the main analyses (eTable 2 in the Supplement).

Discussion

Consistent with a prior study,[4] we found that adolescent social media use was prospectively associated with increased risk of comorbid internalizing and externalizing problems as well as internalizing problems alone. This association remained significant after adjusting for demographics, past alcohol and marijuana use, and, most importantly, a history of mental health problems, which mitigates the possibility that reverse causality explains these findings. In contrast, we did not find an association of social media use with externalizing problems alone. This finding suggests that the association of social media use with comorbid problems occurs primarily because of the association of social media with internalizing problems and the high comorbidity of internalizing and externalizing problems. Unlike a prior study,[4] we found no evidence of moderation by sex, perhaps because of the simplicity of our social media use variable, which could not capture the nature of interactions on social media that may differ by sex.

Numerous mechanisms could account for the association between social media use and internalizing problems. Adolescents who engage in high levels of social media use may experience poorer quality sleep, which may be a mediator on the pathway to internalizing problems.[27] Time spent on social media may increase the risk of experiencing cyberbullying, which has a strong association with depressive symptoms.[28] Social media may also expose adolescents to idealized self-presentations that negatively influence body image and encourage social comparisons.[4] Poor emotion regulation and lack of social interaction may also be associated with social media use and contribute to symptoms of anxiety and depression.[29]

These mechanisms are potentially consistent with the notion that spending less time on social media may contribute to mental health. In fact, the PAFs obtained in our study suggest that if adolescents using social media for more than 30 minutes per day had instead used it for 30 minutes or less, there would have been 9.4% fewer high internalizing problem cases and 7.3% fewer high externalizing problem cases. Of importance, this is not meant to imply that reductions in mental health problems would definitively happen if social media use were reduced or that all social media use is harmful. Instead, these PAFs suggest the potential influence of our findings on the population at a national level assuming a causal effect of social media use and no confounding—both strong assumptions. Future research could improve on our PAF estimates by using data from randomized clinical trials (RCTs).

Our findings must be balanced with the potential benefits of social media use, which include exposure to current events, communication over geographic barriers, and social inclusion for those who may be otherwise excluded in their day-to-day lives (eg, lesbian, bisexual, transgender, queer, and questioning youth).[3] A limitation of our study is that we measured overall time spent on social media; prior studies[30-32] have found that social media use may be positively or negatively associated with mental health depending on which platforms are used and how. Nevertheless, a number of interventions could lead to a reduction in time spent on social media by adolescents, while still allowing for the benefits of such use. The American Academy of Pediatrics has developed a Family Media Use Plan, which can be tailored to specific developmental phases and help parents set reasonable rules for digital media use.[7] Pediatricians and teachers are essential for promoting these plans, as well as helping parents identify problematic social media use in their children.[33] There is also evidence that

Table 3. Estimated Percentages of Adolescent Mental Health Problem Cases Eliminated in Each Counterfactual Scenario of Time Spent on Social Media[a]

Amount of Time Spent on Social Media per Day	Cases, % (95% CI)				
	Internalizing Only	Externalizing Only	Comorbid	All Internalizing	All Externalizing
No more than					
6 h	0.2 (0.2 to 0.2)	0.4 (0.3 to 0.4)	1.2 (1.1 to 1.2)	0.8 (0.8 to 0.9)	0.8 (0.8-0.9)
3 h	2.3 (2.2 to 2.4)	−3.0 (−3.1 to −2.9)	5.5 (5.3 to 5.6)	4.4 (4.3 to 4.5)	1.7 (1.6 to 1.8)
30 min	3.4 (3.3 to 3.5)	0.7 (0.6 to 0.8)	12.4 (12.2 to 12.7)	9.4 (9.2 to 9.5)	7.3 (7.1 to 7.4)
No time spent on social media	12.7 (12.5 to 12.9)	6.9 (6.7 to 7)	22.0 (21.8 to 22.3)	18.9 (18.7 to 19.1)	15.3 (15.2 to 15.5)

[a] The All Internalizing column includes cases of internalizing only and comorbid internalizing and externalizing problems. The All Externalizing column includes cases of externalizing only and comorbid internalizing and externalizing problems.

interventions that promote media literacy, defined as "specific knowledge and skills that can help critical understanding and usage of the media,"[34(p 455)] counteract the harmful association of media use with behavioral health.[34] Also, there is an increasing movement to improve the design of social media platforms; a notable recent example is not displaying the number of "likes" that an Instagram post receives.[35] We believe that technology companies and regulators responsible for social media platforms should consider how these platforms can be designed to minimize risk of mental health problems.

Some researchers have raised concerns that studies on technology use and well-being are limited by publication bias.[36] We believe that this is a legitimate concern given that many studies on this topic, including the present study, are secondary analyses of data not collected for the purpose of studying social media.[36] There appears to be an urgent need for experimental research, specifically a priori registered RCTs that examine interventions designed to reduce social media use. Our study findings suggest a population-level association between social media use and mental health problems, and evidence from RCTs could build on this by examining changes in mental health as a result of changes in social media use. The existing observational study findings and at least 1 RCT in college students[37] appear to be sufficient to justify investment in these trials. In addition, RCTs may be valuable for developing clinical guidelines and informing regulatory policy for social media design.

Limitations

Some limitations of this study should be noted. First, adolescents self-reported the exposure and outcome, which may inflate the observed associations. Second, we measured mental health problems with a self-report questionnaire rather than a diagnostic interview. Third, the validity of self-reported time spent on social media in the PATH study is unknown. Some research suggests that self-reported time on social media may exceed actual use[38]; future studies should consider the use of digital trace data to capture actual time spent using social media.[39] Fourth, social media use continues to change rapidly over time; although our data were collected relatively recently, they may not reflect current trends. Fifth, although our study design mitigates the possibility of reverse causality, some residual confounding from imprecise measurement of prior mental health problems may have been present. Sixth, it remains possible that mental health problems are prospectively associated with social media use, but we could not examine this in the present study because of data limitations. Seventh, it is possible that the observed associations were an artifact of unmeasured confounding. Although we controlled for a number of potential confounders, there may be others, such as physical activity, that we were unable to include because of data limitations.

Conclusions

This study suggests that increased time spent on social media may be a risk factor for internalizing problems in adolescents. Future research should determine whether setting limits on daily social media use, increasing media literacy, and redesigning social media platforms are effective means of reducing the burden of mental health problems in this population.

ARTICLE INFORMATION

Accepted for Publication: June 14, 2019.

Published Online: September 11, 2019. doi:10.1001/jamapsychiatry.2019.2325

Author Affiliations: Department of Mental Health, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland (Riehm, Feder, Tormohlen, Crum, Mojtabai); Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland (Crum); Welch Center for Prevention, Epidemiology and Clinical Research, Johns Hopkins University, Baltimore, Maryland (Crum); Department of Psychiatry and Behavioral Sciences, Johns Hopkins University, Baltimore, Maryland (Crum, Mojtabai); Division of Child and Adolescent Psychiatry, School of Medicine, Johns Hopkins University, Baltimore, Maryland (Young); Department of Behavioral and Community Health, University of Maryland, College Park, College Park (Green); Department of Psychiatry and Behavioral Sciences, Duke University School of Medicine, Durham, North Carolina (Pacek); Washington State Department of Health, Olympia (La Flair).

Author Contributions: Ms Riehm had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Riehm, Feder, Crum, Green, La Flair, Mojtabai.

Acquisition, analysis, or interpretation of data: Riehm, Feder, Tormohlen, Young, Green, Pacek, La Flair.

Drafting of the manuscript: Riehm, Feder, Pacek.

Critical revision of the manuscript for important intellectual content: All authors.

Statistical analysis: Riehm, Feder, Green, Pacek.

Administrative, technical, or material support: Green.

Supervision: Crum, Green, Mojtabai.

Conflict of Interest Disclosures: Dr Young reported receiving grants from the National Institute on Drug Abuse and the Brain and Behavior

Research Foundation during the conduct of the study; receiving grants from Supernus Pharmaceuticals and Psychnostics LLC outside the submitted work; and receiving personal fees from University of Montana's American Indian/Alaska Native Clinical Translational Program. Dr Pacek reported receiving grants from the National Institute on Drug Abuse during the conduct of the study. No other disclosures were reported.

Funding/Support: Ms Riehm was supported by grant 5T32MH014592-39 from the National Institute of Mental Health Psychiatric Epidemiology Training Program (Peter Zandi, principal investigator) and by a doctoral foreign study award from the Canadian Institutes of Health Research. Dr Feder was supported by National Research and Service Award F32DA044699 from the National Institute on Drug Abuse. Ms Tormohlen was supported by grant T32DA007292 (Renee M. Johnson, principal investigator). Dr Young was supported by grant K23DA044288, and Dr Pacek was supported by grant K01DA043413 from the National Institute on Drug Abuse.

Role of the Funder/Sponsor: The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

REFERENCES

1. Pew Research Center. Teens, social media & technology 2018. https://www.pewinternet.org/wp-content/uploads/sites/9/2018/05/PI_2018.05.31_TeensTech_FINAL.pdf. Accessed April 11, 2019.

2. Council on Communications and Media. Media use in school-aged children and adolescents. *Pediatrics*. 2016;138(5):e20162592. doi:10.1542/peds.2016-2592

3. Zink J, Belcher BR, Kechter A, Stone MD, Leventhal AM. Reciprocal associations between screen time and emotional disorder symptoms during adolescence. *Prev Med Rep*. 2019;13:281-288. doi:10.1016/j.pmedr.2019.01.014

4. McCrae N, Gettings S, Purssell E. Social media and depressive symptoms in childhood and adolescence: a systematic review. *Adolesc Res Rev*. 2017;2(4):315-330. doi:10.1007/s40894-017-0053-4

5. Primack BA, Swanier B, Georgiopoulos AM, Land SR, Fine MJ. Association between media use in adolescence and depression in young adulthood: a longitudinal study. *Arch Gen Psychiatry*. 2009;66(2):181-188. doi:10.1001/archgenpsychiatry.2008.532

6. Tsitsika LL, Inkster B. Online social networking sites and mental health research. *Front Psychiatry*. 2015;6:36.

7. Jelenchick LA, Eickhoff JC, Moreno MA. "Facebook depression?" social networking site use and depression in older adolescents. *J Adolesc Health*. 2013;52(1):128-130. doi:10.1016/j.jadohealth.2012.05.008

8. Ra CK, Cho J, Stone MD, et al. Association of digital media use with subsequent symptoms of attention-deficit/hyperactivity disorder among adolescents. *JAMA*. 2018;320(3):255-263. doi:10.1001/jama.2018.8931

9. Galica VL, Vannucci A, Flannery KM, Ohannessian CM. Social media use and conduct problems in emerging adults. *Cyberpsychol Behav Soc Netw*. 2017;20(7):448-452. doi:10.1089/cyber.2017.0068

10. Mojtabai R, Olfson M, Han B. National trends in the prevalence and treatment of depression in adolescents and young adults. *Pediatrics*. 2016;138(6):e20161878. doi:10.1542/peds.2016-1878

11. Twenge JM, Cooper AB, Joiner TE, Duffy ME, Binau SG. Age, period, and cohort trends in mood disorder indicators and suicide-related outcomes in a nationally representative dataset, 2005-2017. *J Abnorm Psychol*. 2019;128(3):185-199. doi:10.1037/abn0000410

12. Hedegaard H, Curtin SC, Warner M. *Suicide Mortality in the United States, 1999-2017*. *NCHS Data Brief No. 330*. Hyattsville, MD: National Center for Health Statistics; 2018.

13. Burstein B, Agostino H, Greenfield B. Suicidal attempts and ideation among children and adolescents in US emergency departments, 2007-2015 [published online April 8, 2019]. *JAMA Pediatr*. doi:10.1001/jamapediatrics.2019.0464

14. Twenge JM, Martin GN, Campbell WK. Decreases in psychological well-being among American adolescents after 2012 and links to screen time during the rise of smartphone technology. *Emotion*. 2018;18(6):765-780. doi:10.1037/emo0000403

15. Hyland A, Ambrose BK, Conway KP, et al. Design and methods of the Population Assessment of Tobacco and Health (PATH) Study. *Tob Control*. 2017;26(4):371-378. doi:10.1136/tobaccocontrol-2016-052934

16. VanderWeele TJ, Jackson JW, Li S. Causal inference and longitudinal data: a case study of religion and mental health. *Soc Psychiatry Psychiatr Epidemiol*. 2016;51(11):1457-1466. doi:10.1007/s00127-016-1281-9

17. US Department of Health and Human Services, National Institutes of Health, National Institute on Drug Abuse, Food and Drug Administration Center for Tobacco Products. *Population Assessment of Tobacco and Health (PATH) Study*. Ann Arbor, MI: PATH; 2017.

18. Dennis ML, Chan YF, Funk RR. Development and validation of the GAIN Short Screener (GSS) for internalizing, externalizing and substance use disorders and crime/violence problems among adolescents and adults. *Am J Addict*. 2006;15(suppl 1):80-91. doi:10.1080/10550490601006055

19. McDonell MG, Comtois KA, Voss WD, Morgan AH, Ries RK. Global Appraisal of Individual Needs Short Screener (GSS): psychometric properties and performance as a screening measure in adolescents. *Am J Drug Alcohol Abuse*. 2009;35(3):157-160. doi:10.1080/00952990902825421

20. Green VR, Conway KP, Silveira ML, et al. Mental health problems and onset of tobacco use among 12- to 24-year-olds in the PATH study. *J Am Acad Child Adolesc Psychiatry*. 2018;57(12):944-954.e4. doi:10.1016/j.jaac.2018.06.029

21. Riehm KE, Young AS, Feder KA, et al. Mental health problems and initiation of e-cigarette and combustible cigarette use. *Pediatrics*. 2019;144(1):e20182935.

22. Schisterman EF, Cole SR, Platt RW. Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology*. 2009;20(4):488-495. doi:10.1097/EDE.0b013e3181a819a1

23. Mantel N. The detection of disease clustering and a generalized regression approach. *Cancer Res*. 1967;27(2):209-220.

24. Greenland S, Drescher K. Maximum likelihood estimation of the attributable fraction from logistic models. *Biometrics*. 1993;49(3):865-872. doi:10.2307/2532206

25. Rückinger S, von Kries R, Toschke AM. An illustration of and programs estimating attributable fractions in large scale surveys considering multiple risk factors. *BMC Med Res Methodol*. 2009;9(1):7. doi:10.1186/1471-2288-9-7

26. Tilling K, Williamson EJ, Spratt M, Sterne JA, Carpenter JR. Appropriate inclusion of interactions was needed to avoid bias in multiple imputation. *J Clin Epidemiol*. 2016;80:107-115. doi:10.1016/j.jclinepi.2016.07.004

27. Li X, Buxton OM, Lee S, Chang AM, Berger LM, Hale L. Sleep mediates the association between adolescent screen time and depressive symptoms. *Sleep Med*. 2019;57:51-60. doi:10.1016/j.sleep.2019.01.029

28. Bottino SMB, Bottino CM, Regina CG, Correia AV, Ribeiro WS. Cyberbullying and adolescent mental health: systematic review. *Cad Saude Publica*. 2015;31(3):463-475. doi:10.1590/0102-311x00036114

29. Hoge E, Bickham D, Cantor J. Digital media, anxiety, and depression in children. *Pediatrics*. 2017;140(suppl 2):S76-S80. doi:10.1542/peds.2016-1758G

30. Seabrook EM, Kern ML, Rickard NS. Social networking sites, depression, and anxiety: a systematic review. *JMIR Ment Health*. 2016;3(4):e50. doi:10.2196/mental.5842

31. Baker DA, Algorta GP. The relationship between online social networking and depression: a systematic review of quantitative studies. *Cyberpsychol Behav Soc Netw*. 2016;19(11):638-648. doi:10.1089/cyber.2016.0206

32. Vannucci V, Johnson A, Villanti AC, Evans WD, Turner M. Patterns of social media use and their relationship to health risks among young adults. *J Adolesc Health*. 2019;64(2):158-164. doi:10.1016/j.jadohealth.2018.06.025

33. Joshi SV, Stubbe D, Li S-TT, Hilty DM. The use of technology by youth: implications for psychiatric educators. *Acad Psychiatry*. 2019;43(1):101-109. doi:10.1007/s40596-018-1007-2

34. Jeong S-H, Cho H, Hwang Y. Media literacy interventions: a meta-analytic review. *J Commun*. 2012;62(3):454-472. doi:10.1111/j.1460-2466.2012.01643.x

35. Kuleff K. Instagram is testing hiding your likes. CNN website. https://www.cnn.com/2019/04/30/tech/instagram-hiding-likes/index.html. Accessed June 12, 2019.

36. Orben A, Przybylski AK. The association between adolescent well-being and digital technology use. *Nat Hum Behav*. 2019;3(2):173-182. doi:10.1038/s41562-018-0506-1

37. Hunt MG, Marx R, Lipson C, et al. No more FOMO: limiting social media decreases loneliness and depression. *J Soc Clin Psychol*. 2018;37(10):751-768. doi:10.1521/jscp.2018.37.10.751

38. Junco R. Comparing actual and self-reported measures of Facebook use. *Comput Human Behav*. 2013;29(3):626-631. doi:10.1016/j.chb.2012.11.007

39. Stier S, Breuer J, Siegers P, et al. Integrating survey data and digital trace data: key issues in developing an emerging field [published online April 24, 2019]. *Soc Sci Comput Rev*. doi:10.1177/0894439319843669

JAMA Netw Open. 2025 May 21;8(5):e2511704. doi: 10.1001/jamanetworkopen.2025.11704

Social Media Use and Depressive Symptoms During Early Adolescence

Jason M Nagata [1,✉], Christopher D Otmar [1], Joan Shim [1], Priyadharshini Balasubramanian [1], Chloe M Cheng [1], Elizabeth J Li [1], Abubakr A A Al-Shoaibi [1], Iris Y Shao [1], Kyle T Ganson [2], Alexander Testa [3], Orsolya Kiss [4], Jinbo He [5], Fiona C Baker [4]

Author information Article notes Copyright and License information
PMCID: PMC12096259 PMID:

See "Toward Defining Probl n JAMA, 10.1001/jama.2025.6113.

Depression

This cohort study of chil valuates person-level associations between tim symptoms over subsequent years.

Key Points

Question

Are there within-person associations between social media use (time) and depressive symptoms across early adolescence?

JAMA Netw Open. 2025 May 21;8(5):e2511704. doi: 10.1001/jamanetworkopen.2025.11704

Social Media Use and Depressive Symptoms During Early Adolescence

Jason M Nagata [1,✉], Christopher D Otmar [1], Joan Shim [1], Priyadharshini Balasubramanian [1], Chloe M Cheng [1], Elizabeth J Li [1], Abubakr A A Al-Shoaibi [1], Iris Y Shao [1], Kyle T Ganson [2], Alexander Testa [3], Orsolya Kiss [4], Jinbo He [5], Fiona C Baker [4]

Author information Article notes Copyright and License information
PMCID: PMC12096259 PMID: 40397441

See "Toward Defining Problematic Media Usage Patterns in Adolescents," in *JAMA*, 10.1001/jama.2025.6113.

This cohort study of children and adolescents aged 9 to 12 years evaluates person-level associations between time spent using social media and depressive symptoms over subsequent years.

Key Points

Question

Are there within-person associations between social media use (time) and depressive symptoms across early adolescence?

Findings

In this cohort study of 11 876 children and adolescents, within-person increases in social media use during early adolescence were prospectively associated with greater depressive symptoms 1 year later, whereas depressive symptoms were not associated with later social media use.

Meaning

The findings suggest that more time spent on social media during early adolescence may contribute to increased depressive symptoms over time.

Abstract

Importance

In 2023, the US Surgeon General issued the Advisory on Social Media and Youth Mental Health, identifying critical research gaps that preclude evidence-based guidance given that most studies of social media and mental health have been cross-sectional rather than longitudinal and have focused on young adults or older adolescents rather than on younger adolescents.

Objective

To evaluate longitudinal associations between social media use (time spent on social media) and depressive symptoms across 4 annual waves spanning a 3-year follow-up period from late childhood to early adolescence.

Design, Setting, and Participants

In this prospective cohort study using data from the Adolescent Brain Cognitive Development Study across 21 study sites from October 2016 to October 2018, children aged 9 to 10 years at baseline were assessed across 4 waves (baseline, year 1, year 2, and year 3), with year-3 follow-up through 2022. Sample sizes varied across waves and measures due to attrition and missing data. Analyses retained all available data at each wave. Data were analyzed from January 2024 to March 2025.

Exposures

Self-reported time spent on social media at baseline to 3-year follow-up.

Main Outcomes and Measures

Reciprocal associations between social media use and depressive symptoms (Child Behavior Checklist) at baseline and at 1, 2, and 3 years of follow-up were assessed using longitudinal, cross-lagged structural equation panel models. Covariates included sex, race and ethnicity, household income, and parental educational level.

Results

At baseline, the sample included 11 876 participants (mean [SD] age, 9.9 [0.6] years), of whom 6196 (52.2%) were male. After adjusting for stable between-person differences and covariates, within-person increases in social media use above the person-level mean were associated with elevated depressive symptoms from year 1 to year 2 (β, 0.07; 95% CI, 0.01-0.12; *P* = .01) and from year 2 to year 3 (β, 0.09; 95% CI, 0.04-0.14; *P* < .001), whereas depressive symptoms were not associated with subsequent social media use at any interval. The final random-intercept cross-lagged panel model demonstrated a good fit (comparative fit index, 0.977; Tucker-Lewis index, 0.968; root mean square error of approximation, 0.031 [90% CI, 0.029-0.033]). Between-person differences in social media use were not associated with depressive symptoms (β, −0.01; 95% CI, −0.04 to 0.02; *P* = .46) after accounting for demographic and family-level factors.

Conclusions and Relevance

In this cohort study of 11 876 children and adolescents, reporting higher than person-level mean social media use in years 1 and 2 after baseline was associated with greater depressive symptoms in the subsequent year. The findings suggest that clinicians should provide anticipatory guidance regarding social media use for young adolescents and their parents.

Introduction

Social media use among adolescents has risen sharply in recent years, raising concerns about its impact on mental health.[1] In 2021, 42% of adolescents reported persistent feelings of sadness or

hopelessness, an increase of 50% from 2011.[2] Although correlations between social media use and depressive symptoms have been previously identified, the directionality of this relationship remains unclear.[3,4,5,6,7,8,9] In 2023, the US Surgeon General issued the Advisory on Social Media and Youth Mental Health,[10] calling for longitudinal research, as most prior studies have been cross-sectional and were therefore unable to determine temporality, directionality, or within-person changes.[3,4,5,6,7,8,9] Disentangling whether social media use contributes to or is a reflection of preexisting distress is critical for guiding evidence-based interventions and policy decisions.

Building on these concerns, the Differential Susceptibility to Media Effects Model (DSMM)[11] provides a guiding framework for understanding the relationship between social media use and adolescent mental health. The DSMM posits that media effects are not uniform but depend on dispositional, developmental, and sociocultural factors, which in adolescence may include heightened cognitive and emotional reactivity.[12,13] These sensitivities make adolescence a critical period of vulnerability during which social media exposure may have lasting implications for mental health.[14] Social media use may also play a bidirectional role; it can influence future mood states while also being shaped by preexisting depressive symptoms, potentially creating reinforcing cycles of use and distress.[15] By applying the DSMM to examine within-person changes over time,[16] this study aimed to identify whether there are bidirectional associations between social media use and depressive symptoms in early adolescence. Of note, the few existing longitudinal studies on social media and mental health in adolescents have reported mixed findings,[17] and bidirectional relationships remain understudied.

To address this gap, we leveraged data from the Adolescent Brain Cognitive Development (ABCD) Study, an ongoing national prospective cohort study that tracks participants over multiple time points.[18,19] This design enabled us to explore individual trajectories and within-person variability in the association between social media use and depressive symptoms. The design also accounted for autoregressive effects and allowed us to explore the stability of social media use and depressive symptoms over time. We hypothesized that social media use and depressive symptoms in early adolescence would exhibit bidirectional within-person associations over time.

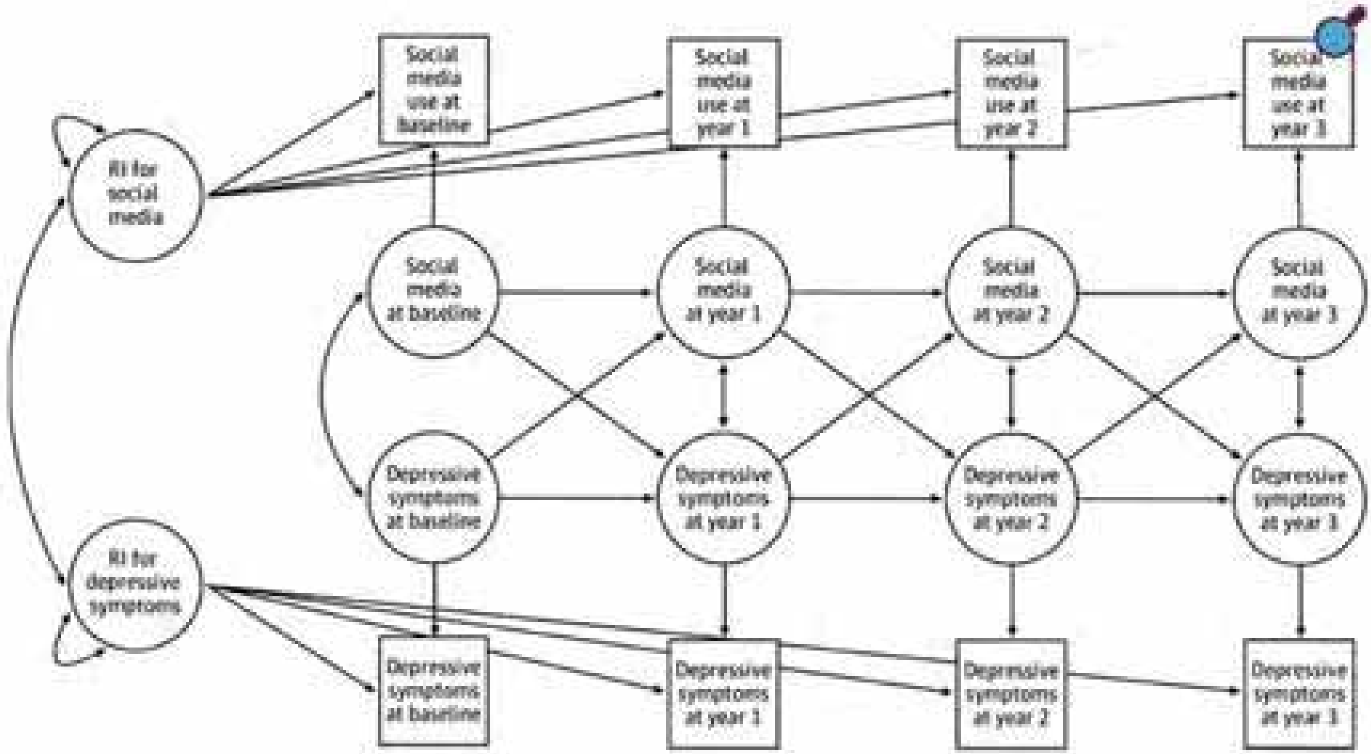
Methods

Study Population

In this cohort study, we conducted analyses of data from baseline to year-3 follow-up of the ABCD Study (5.1 release). The ABCD Study is the largest longitudinal study of adolescent health, brain,

and cognitive development in the US. It recruited children aged 9 to 10 years from 21 sites from October 2016 October 2018 (baseline). Participants were assessed across 4 waves (baseline, year 1, year 2, and year 3), with year-3 follow-up through 2022. Specifically, the repeated assessments allowed for a clearer examination of potential directionality—whether changes in social media use preceded shifts in depressive symptoms or vice versa (Figure 1). Sample sizes varied across waves and measures due to attrition and missing data. Analyses retained all available data at each wave. The ABCD Study sample, recruitment, protocol, and measures were reported previously[20] and are further described in the eMethods in Supplement 1. Centralized institutional review board approval for the ABCD Study was received from the University of California, San Diego; written assent was obtained from the study participants, and written consent was obtained from their parents or guardians. The current study was a secondary analysis of deidentified ABCD Study and thus did not require additional approval or assent and consent. The current study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline for cohort studies.

Figure 1. Conceptual Random-Intercept (RI) Cross-Lagged Panel Model of Social Media Use and Depressive Symptoms.



Open in a new tab

Measures

Social Media Use

Social media use was defined for this analysis as time spent on social media daily and was assessed using the ABCD Youth Screen Time Survey administered each year.[21] Adolescents responded to questions about the number of hours and minutes per weekday and weekend day they spent engaging with social media. The total time spent on social media was calculated as the weighted sum: [(weekday mean × 5) + (weekend mean × 2)]/7. Weighted mean social media use was reported as a continuous variable, representing the mean daily time spent (in hours).

Depressive Symptoms

Depressive symptoms were assessed using the validated Child Behavior Checklist (CBCL) depressive problems score (from the *Diagnostic and Statistical Manual of Mental Disorders*–oriented scales) each year as reported by the caregiver.[22] The raw score (as opposed to the *t* score) was chosen to capture the full distribution of symptom severity in a community-based sample and to avoid the potential loss of within-person variability over time that may occur when standardizing scores according to age- and gender-based norms. Each raw score reflects the sum of items within the depression-related subscale, with higher values indicating more severe symptoms.

Covariates

Several covariates were included to account for demographic and contextual factors that may be associated with social media use or depressive symptoms, including sex (female, male), race and ethnicity (ascertained by parent or caregiver report; categories were Asian, Black, Hispanic or Latino, Native American, White, or other [no defined groups, although write-ins were allowed]), household income (<$25 000, $25 000-$49 999, $50 000-$74 999, $75 000-$99 999, $100 000-$199 999, and ≥$200 000), and highest parental educational level (high school or less vs college or more). The number of adverse childhood experiences,[23] the parental monitoring scale,[24] family conflict (conflict subscale of the Family Environment Scale[25]), and the study site were also included as covariates (the eMethods in Supplement 1 give further details). All covariates were chosen based on prior literature suggesting their relevance to mental health and digital media behaviors.[26]

Statistical Analysis

To examine the reciprocal associations between social media use and depressive symptoms over 4 time points, we fit several longitudinal structural equation models.[27] We first estimated a traditional cross-lagged panel model (CLPM) and then compared it with random-intercept CLPMs (RI-CLPMs). The RI-CLPM is an extension of the traditional CLPM that explicitly separates stable between-person differences from within-person fluctuations over time.[28] By introducing latent random intercepts, the RI-CLPM ensures that each individual's trait-like baseline is accounted for, allowing the cross-lagged paths to focus on how temporary (state-like) deviations in one variable are associated with subsequent deviations in another. This approach is particularly valuable when examining processes that may be confounded by persistent individual differences because it isolates the time-varying relationships within persons.[16] In doing so, the RI-CLPM clarifies whether an association is driven by people who have generally high scores on both constructs and by within-person changes that unfold across measurement occasions. It also enables the

estimation of autoregressive and cross-lagged parameters that reflect carryover and spillover effects independent of any overall rank-order stability between participants.

All models were estimated using maximum likelihood with robust SEs (MLR) in the lavaan package[29] in R, version 4.4.2 (R Project for Statistical Computing). MLR accounts for nonnormality and provides robust SEs and a scaled test statistic. Full information maximum likelihood was used to handle missing data on the outcome measures (the eMethods in Supplement 1 provide more detail on missing data). We evaluated model fit using the comparative fit index (CFI), Tucker-Lewis index (TLI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR). We relied on fit criteria indicating good to excellent model fit[30] (eg, CFI ≥0.95, TLI ≥0.95, RMSEA ≤0.06, and SRMR ≤0.08). In accordance with the benchmarks of Orth et al,[31] cross-lagged effect sizes were interpreted as small ($\beta = 0.03$), medium ($\beta = 0.07$), and large ($\beta = 0.12$). Two-sided $P < .05$ was considered significant. Analyses were performed from January 2024 to March 2025.

Results

Sociodemographic characteristics of the analytic sample after excluding participants with missing baseline data for age or sex assigned at birth (N = 11 876) are presented in Table 1. A total of 5680 participants (47.8%) were female, 6196 (52.2%) were male, and the mean (SD) age at baseline was 9.9 (0.6) years. In all, 709 participants (6.0%) were Asian; 2392 (20.1%), Black; 2027 (17.0%), Hispanic or Latino; 410 (3.4%), Native American; 6166 (51.9%), White; and 171 (1.4%), other race and ethnicity. Table 2 shows descriptive statistics and 0-order correlations for the untransformed social media use variables and the CBCL depression raw scores across waves; descriptive indices (mean, SD, and range) suggested an overall increase in mean daily social media use from baseline to year 3 and a modest increase in mean depressive symptom scores.

Table 1. Sociodemographic Characteristics of ABCD Study Participants at Baseline.

Characteristic	Participants, No. (%) (N = 11 876)
Sex	
Female	5680 (47.8)
Male	6196 (52.2)
Race and ethnicity[a]	
Asian	709 (6.0)
Black	2392 (20.1)
Hispanic or Latino	2027 (17.0)
Native American	410 (3.4)
White	6166 (51.9)
Other	171 (1.4)
Household income, $	
≤24 999	1633 (13.8)
25 000-49 999	1588 (13.4)
50 000-74 999	1498 (12.6)
75 000-99 999	1570 (13.2)
100 000-199 999	3311 (27.9)
≥200 000	1250 (10.5)
Parent's highest educational level	
College or more	2039 (17.2)
High school or less	9799 (82.8)

Open in a new tab

Abbreviation: ABCD, Adolescent Brain Cognitive Development.

[a] Reported by the parent and/or caregiver of the participant. The "other" category had no specific racial or ethnic groups defined, although write-ins were allowed.

Table 2. Descriptive Statistics and 0-Order Correlations of Key Variables of Social Media Use, Depressive Symptoms, and Covariates Across Baseline and 3 Follow-Up Years.

Variable	*r*								
	1	2	3	4	5	6	7	8	9
1. Social media use, baseline									
2. Social media use, year 1	0.37								
3. Social media use, year 2	0.28	0.44							
4. Social media use, year 3	0.24	0.38	0.56						
5. Depressive symptoms, baseline	0.03	0.03	0.00	0.00					
6. Depressive symptoms, year 1	0.03	0.02	0.01	0.01	0.64				
7. Depressive symptoms, year 2	0.03	0.04	0.05	0.04	0.55	0.62			
8. Depressive	0.03	0.05	0.05	0.06	0.47	0.55	0.63		

Variable	r								
	1	2	3	4	5	6	7	8	9
symptoms, year 3									
9. Adverse childhood experiences	0.08	0.06	0.06	0.07	0.20	0.18	0.16	0.16	
10. Parental media monitoring	−0.01	0.02	0.03	0.02	−0.09	−0.10	−0.08	−0.07	−0.14
11. Family conflict	0.09	0.04	0.04	0.04	0.11	0.11	0.08	0.07	0.22

Open in a new tab

[a] Variables 1 through 4 are hours per day and were transformed in the final analytic model because the first 2 years were reported in ordinal categories while the last 2 were reported in continuous hours (0-24). Variables 6 through 11 are scores (details of scores are given in the Depressive Symptoms subsection of the Methods section and the eMethods in Supplement 1).

To examine the reciprocal associations among social media use and depressive symptoms over time and to concurrently partition stable, between-person differences from within-person fluctuations, we estimated 3 alternative longitudinal structural equation models. Model comparisons were conducted in a sequential manner and based on standard global fit indices.[30] Following Chen's[32] recommended cutoff criteria for fit indices (eg, CFI ≥0.95, RMSEA ≤0.06, change in CFI ≤0.01, and change in RMSEA ≤0.015), we evaluated whether each model demonstrated good fit and whether differences in fit between nested models were meaningful. First, we estimated the traditional CLPM, which does not partition stable between-person variance and yielded a poor fit (CFI, 0.917; TLI, 0.807; RMSEA, 0.132 [90% CI, 0.127-0.137]; SRMR, 0.065).

Next, we evaluated the constrained RI-CLPM, which incorporates latent random intercepts to capture trait-like differences and imposes equality constraints on the cross-lagged paths across

time. The constrained model provided markedly improved fit indices (CFI, 0.966; TLI, 0.957; RMSEA, 0.036 [90% CI, 0.034-0.038]; SRMR, 0.027). Compared with the traditional CLPM, the constrained RI-CLPM showed improved model fit (change in CFI, −0.049; change in RMSEA, 0.096). In addition, we estimated the unconstrained RI-CLPM, which allowed the cross-lagged parameters to vary freely across time. This model showed further improvement in fit (CFI, 0.977; TLI, 0.968; RMSEA, 0.031 [90% CI, 0.029-0.033]; SRMR, 0.022). Compared with the constrained RI-CLPM, the unconstrained version improved model fit modestly (change in CFI, 0.011; change in RMSEA, −0.005). On the basis of these comparisons and given the theoretical merit of allowing time-specific associations,[14] we retained the unconstrained RI-CLPM as our final model (eTable 1 in Supplement 1). In the unconstrained RI-CLPM, social media use and depressive symptoms were modeled across the 4 waves (baseline, year 1, year 2, and year 3), controlling for between-person variance and the covariates listed in the Methods.

Between-Person Association

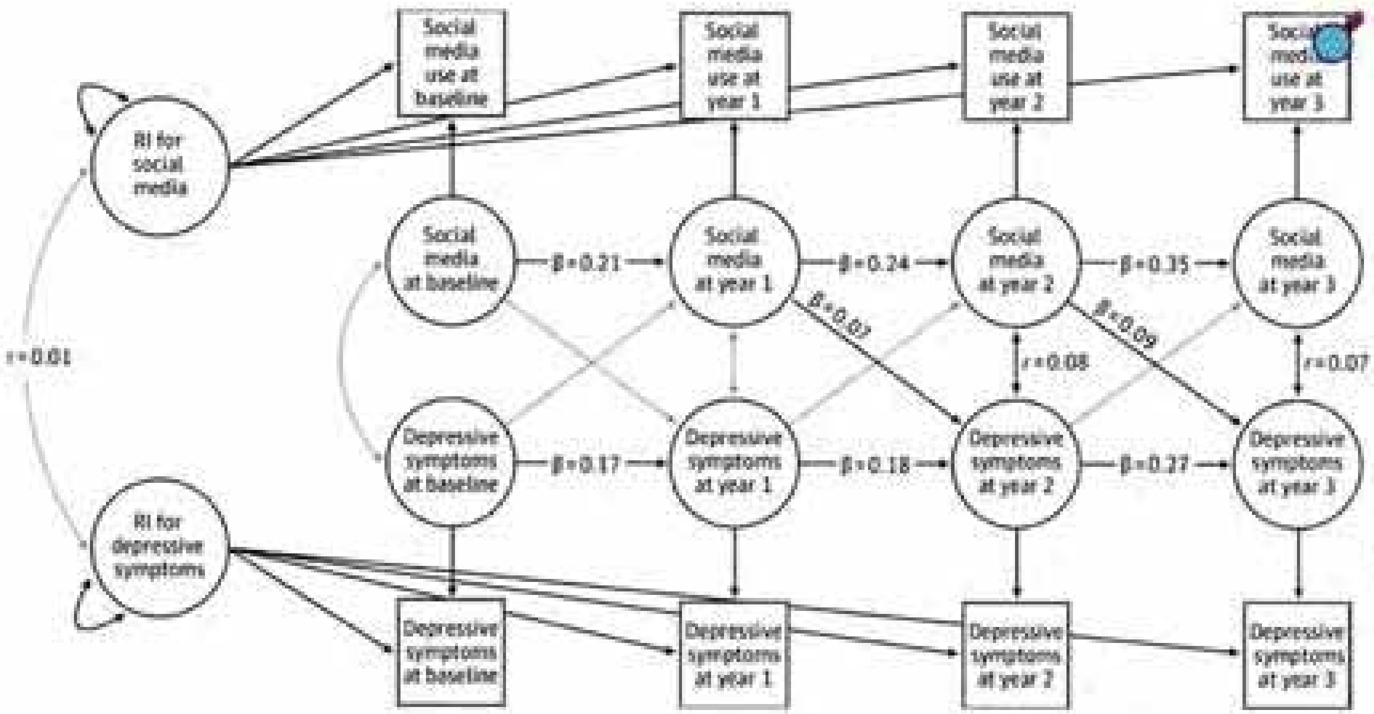
eTable 2 in Supplement 1 displays the standardized estimates for fixed covariates of both social media use and depressive symptoms within the final model. There were no between-person associations between depressive symptoms and social media use, as evidenced by the covariance between the latent random intercepts (β, −0.01; 95% CI, −0.04 to 0.02; $P = .46$). This suggests that adolescents with consistently high (or low) social media use were not necessarily the same adolescents with consistently high (or low) depressive symptoms after accounting for demographic and familial factors and within-person estimates.

Within-Person Associations

The time-varying within-person variables were modeled by estimating both autoregressive and cross-lagged paths among the residuals of the observed indicators (Figure 2 and eTable 3 in Supplement 1). Autoregressive effect sizes were significant for both constructs. Social media use exhibited strong temporal continuity, with an autoregressive coefficient (β) of 0.21 (95% CI, 0.16-0.25; $P < .001$) from baseline to year 1, 0.24 (95% CI, 0.20-0.27; $P < .001$) from year 1 to year 2, and 0.35 (95% CI, 0.32-0.38; $P < .001$) from year 2 to year 3. Depressive symptoms also showed stability, with autoregressive coefficients of 0.17 (95% CI, 0.14-0.20; $P < .001$) from baseline to year 1, 0.18 (95% CI, 0.14-0.21; $P < .001$) from year 1 to year 2, and 0.27 (95% CI, 0.24-0.31; $P < .001$) from year 2 to year 3. Within-wave residual covariance was modeled to assess contemporaneous associations. At year 3, the standardized residual covariance between social media use and depressive symptoms was 0.07 (95% CI, 0.03-0.10; $P < .001$), meaning that individuals reporting

social media use higher than the person-level mean at that wave also tended to have depressive symptoms higher than the person-level mean.

Figure 2. Results of Random-Intercept (RI) Cross-Lagged Panel Model of Social Media Use and Depressive Symptoms in the Adolescent Brain Cognitive Development Study.



Open in a new tab

Black lines represent statistically significant cross-lagged and autoregressive paths and gray lines, nonsignificant paths.

The cross-lagged paths, reflecting how deviations from an individual's expected level on one variable are associated with subsequent deviations on the other, revealed no association between depressive symptoms at baseline and social media use in year 1 (β, 0.00; 95% CI, −0.01 to 0.02; P =.46) or between social media use at baseline and depressive symptoms in year 1 (β, 0.03; 95% CI, −0.02 to 0.09; P =.19). However, significant cross-lagged associations emerged in subsequent waves. Specifically, from year 1 to year 2, social media use higher than the person-level mean in year 1 was associated with greater depressive symptoms in year 2 (β, 0.07; 95% CI, 0.01-0.12; P =.01; medium effect size). This pattern continued from year 2 to year 3, in which social media use higher

than the person-level mean in year 2 was positively associated with greater depressive symptoms in year 3 (β, 0.09; 95% CI, 0.04-0.14; $P < .001$; medium effect size). There were no cross-lagged associations between depressive symptoms and later social media use at either interval (eg, year 1 depressive symptoms and year 2 social media use: β, 0.00 [95% CI, −0.007 to 0.01]; $P = .65$; year 2 depressive symptoms and year 3 social media use: (β, 0.00 [95% CI, −0.003 to 0.02]; $P = .18$).

Discussion

In this cohort study of children and adolescents aged 9 to 12 years in the US, we found that there was a longitudinal association between increases in social media use and subsequent depressive symptoms at the within-person level. These findings provide initial evidence of temporal ordering and could suggest that social media use is a potential contributing factor to adolescent depressive symptoms rather than merely a correlate or consequence of such symptoms. Our findings are consistent with prior studies that have found associations between social media use and depression.[33,34,35] In addition, a meta-analysis of 21 cross-sectional and 5 longitudinal studies found that there was a linear dose-response association between social media use and depression, further suggesting that social media use may be a risk factor for depression.[36]

Only a limited number of studies have assessed bidirectional longitudinal associations between social media use and depression in adolescents, with mixed findings.[37,38,39] One Australian study of individuals aged 10 to 17 years used the RI-CLPM to analyze data from 2013 to 2015 and found no significant cross-lagged associations between social media use and depression.[38] A Dutch study of data from 2016 to 2018 (mean [SD] participant age of 13.1 [0.8] years) using the RI-CLPM found a unidirectional association between problematic social media use and decreased mental health a year later but not vice versa.[39] However, that study found no longitudinal associations between social media use intensity (eg, frequency of viewing, messaging) and mental health in either direction.[39] In contrast, our study found that more time spent on social media, conceptually close to the Dutch study's measure of intensity, was associated with later depressive symptoms. Potential reasons for the differing findings could include variations in the periods (adolescent social media use has increased significantly in the past 15 years), age ranges (the analytic cohort in our study was limited to individuals aged 9 to 12 years), and country (US, Australia, and the Netherlands).

These findings can be interpreted within the context of the DSMM,[11] which posits that some adolescents may be more susceptible to negative media effects due to dispositional (eg, personality, self-esteem), developmental (eg, age), and social-contextual factors (eg, family conflict). Differential susceptibilities may also explain why some social media may be beneficial

for certain individuals while detrimental to others. In later waves of our study (particularly year 3), adolescents who reported social media use higher than the person-level mean also showed depressive symptoms higher than the person-level mean in the same wave. The contemporaneous associations suggest that immediate factors (eg, negative peer social interactions, family conflict) may coincide with or amplify concurrent distress.

The present study's findings have implications for clinical practice and health policy. When interpreting our findings within the context of the DSMM,[11] interventions targeted at addressing developmental and social-contextual factors that may be associated with the negative effects of social media among adolescents could be considered. In particular, given that age is a likely developmental factor associated with these negative outcomes, early detection of and intervention for social media use are important. Furthermore, although our findings suggest that there is a unidirectional association between social media use and depression, with increases in social media use associated with depressive symptoms through subsequent years, prior research on adolescents, especially those with depressive symptoms, has shown that they can shift from maladaptive to more positive patterns of social media use when they become aware of its impact on their mood.[40] Specifically, qualitative interviews with treatment-seeking adolescents with depression revealed that over time, many adjusted their social media behaviors, reducing stress-related posting, avoiding triggering content, and using social media more intentionally to connect with supportive peers.[41] Interventions that promote mindful, purpose-driven social media use, such as encouraging adolescents to prioritize social connection,[41] may help mitigate negative outcomes and support better mental health.

Clinicians should consider inquiring about social media use among children and adolescents, particularly those younger than the recommended age limits (the minimum age requirement for most social media platforms is 13 years), and providing anticipatory guidance as needed. Professional organizations, such as the American Academy of Pediatrics, could refine guidelines on social media use and emphasize the importance of family media plans and intentional social media use.

Strengths and Limitations

The strengths and limitations of this study should be noted. Our study adds to knowledge in the field of adolescent health and communication science by examining longitudinal associations between social media use and depressive symptoms over 4 years, whereas most previous studies were cross-sectional. In addition, strengths include the analysis of a large, demographically diverse, contemporary sample of children and adolescents in the US. Limitations include the

observational design of the study, leading to susceptibility to residual and unmeasured confounders despite adjustment for potential confounders, as well as reporting, recall, and social desirability bias.[42]

Conclusions

In this cohort study of participants enrolled in the ABCD Study at age 9 or 10 years, higher person-level social media use in years 1 and 2 was associated with greater depressive symptoms in years 2 and 3. These findings suggest that more time spent on social media during early adolescence may contribute to increased depressive symptoms over time. Future studies could examine whether social media use is linked to heightened depressive symptoms by examining short-term shifts in cognitive (eg, negative self-talk, social comparison, or rumination) and excitative (eg, physiological arousal or stress) states, both of which are outlined as mediators in the DSMM.[11] Given that these states may fluctuate over days, weeks, or seasons, more intensive within-person designs (eg, daily diaries, ecological momentary assessment, and passive mobile sensing via smartphone) may offer a more precise understanding of these processes compared with annual assessments. Future research should also continue to examine the prospective relationships between social media and mental health outcomes as the ABCD Study cohort ages to middle and late adolescence as well as aim to examine the mechanisms underlying these associations.

Supplement 1.

eMethods. Study Design, Sample, Missing Data, Model Comparison and Construction, and Fixed Covariates

eTable 1. Model Fit Indices for Traditional, Constrained, and Unconstrained Random-Intercept Cross-Lagged Panel Models

eTable 2. Fixed Covariates in the Unconstrained Random-Intercept Cross-Lagged Panel Model for Social Media Use and Depressive Symptoms

eTable 3. Estimated Residual Covariances, Autoregressive Paths, and Cross-Lagged Associations Between Social Media Use and Depressive Symptoms in the Unconstrained Random-Intercept Cross-Lagged Panel Model

eReferences

jamanetwopen-e2511704-s001.pdf (242KB, pdf)

Supplement 2.

Data Sharing Statement

jamanetwopen-e2511704-s002.pdf (17.5KB, pdf)

References

1. Jaycox LH, Murphy ER, Zehr JL, Pearson JL, Avenevoli S. Social media and suicide risk in youth. JAMA Netw Open. 2024;7(10):e2441499. doi: 10.1001/jamanetworkopen.2024.41499 [DOI] [PubMed] [Google Scholar]

2. Youth Risk Behavior Surveillance System. *Youth Risk Behavior Survey Data Summary & Trends Report 2013-2023*. Centers for Disease Control and Prevention. August 6, 2024. Accessed March 2, 2023. https://www.cdc.gov/yrbs/dstr/index.html

3. Schou Andreassen C, Billieux J, Griffiths MD, et al. The relationship between addictive use of social media and video games and symptoms of psychiatric disorders: a large-scale cross-sectional study. Psychol Addict Behav. 2016;30(2):252-262. doi: 10.1037/adb0000160 [DOI] [PubMed] [Google Scholar]

4. Yoo HJ, Cho SC, Ha J, et al. Attention deficit hyperactivity symptoms and internet addiction. Psychiatry Clin Neurosci. 2004;58(5):487-494. doi: 10.1111/j.1440-1819.2004.01290.x [DOI] [PubMed] [Google Scholar]

5. Yen JY, Ko CH, Yen CF, Wu HY, Yang MJ. The comorbid psychiatric symptoms of internet addiction: attention deficit and hyperactivity disorder (ADHD), depression, social phobia, and hostility. J Adolesc Health. 2007;41(1):93-98. doi: 10.1016/j.jadohealth.2007.02.002 [DOI] [PubMed] [Google Scholar]

6. Lakkunarajah S, Adams K, Pan AY, Liegl M, Sadhir M. A trying time: problematic internet use (PIU) and its association with depression and anxiety during the COVID-19 pandemic. Child Adolesc Psychiatry Ment Health. 2022;16(1):49. doi: 10.1186/s13034-022-00479-6 [DOI] [PMC free article] [PubMed] [Google Scholar]

7. Buja A, Gallimberti L, Chindamo S, et al. Problematic social networking site usage and substance use by young adolescents. BMC Pediatr. 2018;18(1):367. doi: 10.1186/s12887-018-1316-3 [DOI] [PMC free article] [PubMed] [Google Scholar]

8. Mathews CL, Morrell HER, Molle JE. Video game addiction, ADHD symptomatology, and video game reinforcement. Am J Drug Alcohol Abuse. 2019;45(1):67-76. doi: 10.1080/00952990.2018.1472269 [DOI] [PubMed] [Google Scholar]

9. Haug S, Castro RP, Kwon M, Filler A, Kowatsch T, Schaub MP. Smartphone use and smartphone addiction among young people in Switzerland. *J Behav Addict*. 4(4):299-307. doi: 10.1556/2006.4.2015.037 [DOI] [PMC free article] [PubMed]

10. Office of the US Surgeon General. Social media and youth mental health: the US Surgeon General's Advisory. 2023. Accessed November 6, 2023. https://www.hhs.gov/sites/default/files/sg-youth-mental-health-social-media-advisory.pdf

11. Valkenburg PM, Peter J. The Differential Susceptibility to Media Effects Model. J Commun. 2013;63(2):221-243. doi: 10.1111/jcom.12024 [DOI] [Google Scholar]

12. Fuhrmann D, Knoll LJ, Blakemore SJ. Adolescence as a sensitive period of brain development. Trends Cogn Sci. 2015;19(10):558-566. doi: 10.1016/j.tics.2015.07.008 [DOI] [PubMed] [Google Scholar]

13. Sisk LM, Gee DG. Stress and adolescence: vulnerability and opportunity during a sensitive window of development. Curr Opin Psychol. 2022;44:286-292. doi: 10.1016/j.copsyc.2021.10.005 [DOI] [PMC free article] [PubMed] [Google Scholar]

14. Orben A, Przybylski AK, Blakemore SJ, Kievit RA. Windows of developmental sensitivity to social media. Nat Commun. 2022;13(1):1649. doi: 10.1038/s41467-022-29296-3 [DOI] [PMC free article] [PubMed] [Google Scholar]

15. Flannery JS, Maza MT, Kilic Z, Telzer EH. Cascading bidirectional influences of digital media use and mental health in adolescence. Adv Child Dev Behav. 2023;64:255-287. doi: 10.1016/bs.acdb.2022.10.003 [DOI] [PubMed] [Google Scholar]

16. Hamaker EL. Why researchers should think "within-person": a paradigmatic rationale. In: Mehl MR, Conner TS, eds. Handbook of Research Methods for Studying Daily Life. Guilford Press; 2012:43-61. [Google Scholar]

17. Lopes LS, Valentini JP, Monteiro TH, et al. Problematic social media use and its relationship with depression or anxiety: a systematic review. Cyberpsychol Behav Soc Netw. 2022;25(11):691-702. doi: 10.1089/cyber.2021.0300 [DOI] [PubMed] [Google Scholar]

18. Curran PJ, Bauer DJ. The disaggregation of within-person and between-person effects in longitudinal models of change. Annu Rev Psychol. 2011;62:583-619. doi: 10.1146/annurev.psych.093008.100356 [DOI] [PMC free article] [PubMed] [Google Scholar]

19. Wang LP, Maxwell SE. On disaggregating between-person and within-person effects with longitudinal data using multilevel models. Psychol Methods. 2015;20(1):63-83. doi: 10.1037/met0000030 [DOI] [PubMed] [Google Scholar]

20. Garavan H, Bartsch H, Conway K, et al. Recruiting the ABCD sample: design considerations and procedures. Dev Cogn Neurosci. 2018;32:16-22. doi: 10.1016/j.dcn.2018.04.004 [DOI] [PMC free article] [PubMed] [Google Scholar]

21. Bagot KS, Tomko RL, Marshall AT, et al. Youth screen use in the ABCD Study. Dev Cogn Neurosci. 2022;57:101150. doi: 10.1016/j.dcn.2022.101150 [DOI] [PMC free article] [PubMed] [Google Scholar]

22. Achenbach TM. The Child Behavior Checklist and related instruments. In: Maruish ME, ed. *The Use of Psychological Testing for Treatment Planning and Outcomes Assessment*. 2nd ed. Lawrence Erlbaum Associates; 1999:429-466. [Google Scholar]

23. Nagata JM, Trompeter N, Singh G, et al. Adverse childhood experiences and early adolescent cyberbullying in the United States. J Adolesc. 2023;95(3):609-616. doi: 10.1002/jad.12124 [DOI] [PMC free article] [PubMed] [Google Scholar]

24. Karoly HC, Callahan T, Schmiege SJ, Ewing SW. Evaluating the Hispanic paradox in the context of adolescent risky sexual behavior: the role of parent monitoring. J Pediatr Psychol. 2016;41(4):429-440. doi: 10.1093/jpepsy/jsv039 [DOI] [PMC free article] [PubMed] [Google Scholar]

25. Moos RH, Moos BS. A typology of family social environments. Fam Process. 1976;15(4):357-371. doi: 10.1111/j.1545-5300.1976.00357.x [DOI] [PubMed] [Google Scholar]

26. Odgers CL, Jensen MR. Annual research review: adolescent mental health in the digital age: facts, fears, and future directions. J Child Psychol Psychiatry. 2020;61(3):336-348. doi: 10.1111/jcpp.13190 [DOI] [PMC free article] [PubMed] [Google Scholar]

27. Usami S. On the differences between general cross-lagged panel model and random-intercept cross-lagged panel model: interpretation of cross-lagged parameters and model choice. Struct Equ Modeling. 2021;28(3):331-344. doi: 10.1080/10705511.2020.1821690 [DOI] [Google Scholar]

28. Hamaker EL, Kuiper RM, Grasman RPPP. A critique of the cross-lagged panel model. Psychol Methods. 2015;20(1):102-116. doi: 10.1037/a0038889 [DOI] [PubMed] [Google Scholar]

29. Rosseel Y. Lavaan: an R package for structural equation modeling. J Stat Softw. 2012;48(2):1-36. doi: 10.18637/jss.v048.i02 [DOI] [Google Scholar]

30. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Modeling. 1999;6(1):1-55. doi: 10.1080/10705519909540118 [DOI] [Google Scholar]

31. Orth U, Meier LL, Bühler JL, et al. Effect size guidelines for cross-lagged effects. Psychol Methods. 2024;29(2):421-433. doi: 10.1037/met0000499 [DOI] [PubMed] [Google Scholar]

32. Chen FF. Sensitivity of goodness of fit indexes to lack of measurement invariance. Struct Equ Modeling. 2007;14(3):464-504. doi: 10.1080/10705510701301834 [DOI] [Google Scholar]

33. Weigle PE, Shafi RMA. Social media and youth mental health. Curr Psychiatry Rep. 2024;26(1):1-8. doi: 10.1007/s11920-023-01478-w [DOI] [PubMed] [Google Scholar]

34. Woods HC, Scott H. #Sleepyteens: social media use in adolescence is associated with poor sleep quality, anxiety, depression and low self-esteem. J Adolesc. 2016;51:41-49. doi: 10.1016/j.adolescence.2016.05.008 [DOI] [PubMed] [Google Scholar]

35. Twenge JM, Joiner TE, Rogers ML, Martin GN. Increases in depressive symptoms, suicide-related outcomes, and suicide rates among US adolescents after 2010 and links to increased new media screen time. Clin Psychol Sci. 2018;6(1):3-17. doi: 10.1177/2167702617723376 [DOI] [Google Scholar]

36. Liu M, Kamper-DeMarco KE, Zhang J, Xiao J, Dong D, Xue P. Time spent on social media and risk of depression in adolescents: a dose-response meta-analysis. Int J Environ Res Public Health. 2022;19(9):5164. doi: 10.3390/ijerph19095164 [DOI] [PMC free article] [PubMed] [Google Scholar]

37. Romer D, Bagdasarov Z, More E. Older versus newer media and the well-being of United States youth: results from a national longitudinal panel. J Adolesc Health. 2013;52(5):613-619. doi: 10.1016/j.jadohealth.2012.11.012 [DOI] [PubMed] [Google Scholar]

38. Houghton S, Lawrence D, Hunter SC, et al. Reciprocal relationships between trajectories of depressive symptoms and screen media use during adolescence. J Youth Adolesc. 2018;47(11):2453-2467. doi: 10.1007/s10964-018-0901-y [DOI] [PMC free article] [PubMed] [Google Scholar]

39. Boer M, Stevens GWJM, Finkenauer C, de Looze ME, van den Eijnden RJJM. Social media use intensity, social media use problems, and mental health among adolescents: investigating directionality and mediating processes. Comput Human Behav. 2021;116:106645. doi: 10.1016/j.chb.2020.106645 [DOI] [Google Scholar]

40. Radovic A, Gmelin T, Stein BD, Miller E. Depressed adolescents' positive and negative use of social media. J Adolesc. 2017;55:5-15. doi: 10.1016/j.adolescence.2016.12.002 [DOI] [PMC free article] [PubMed] [Google Scholar]

41. Holt-Lunstad J. Social connection as a public health issue: the evidence and a systemic framework for prioritizing the "social" in social determinants of health. Annu Rev Public Health. 2022;43:193-213. doi: 10.1146/annurev-publhealth-052020-110732 [DOI] [PubMed] [Google Scholar]

42. Van de Mortel TF. Faking it: social desirability response bias in self-report research. Aust J Adv Nurs. 2008;25(4):40-48. [Google Scholar]

Associated Data

This section collects any data citations, data availability statements, or supplementary materials included in this article.

Supplementary Materials

Supplement 1.

eMethods. Study Design, Sample, Missing Data, Model Comparison and Construction, and Fixed Covariates

eTable 1. Model Fit Indices for Traditional, Constrained, and Unconstrained Random-Intercept Cross-Lagged Panel Models

eTable 2. Fixed Covariates in the Unconstrained Random-Intercept Cross-Lagged Panel Model for Social Media Use and Depressive Symptoms

eTable 3. Estimated Residual Covariances, Autoregressive Paths, and Cross-Lagged Associations Between Social Media Use and Depressive Symptoms in the Unconstrained Random-Intercept Cross-Lagged Panel Model

eReferences

jamanetwopen-e2511704-s001.pdf (242KB, pdf)

Supplement 2.

Data Sharing Statement

jamanetwopen-e2511704-s002.pdf (17.5KB, pdf)

JAMA | Original Investigation

Addictive Screen Use Trajectories and Suicidal Behaviors, Suicidal Ideation, and Mental Health in US Youths

Yunyu Xiao, PhD; Yuan Meng, PhD; Timothy T. Brown, PhD; Katherine M. Keyes, PhD; J. John Mann, MD

Editorial

Supplemental content

IMPORTANCE Increasing child and adolescent use of social media, video games, and mobile phones has raised concerns about potential links to youth mental health problems. Prior research has largely focused on total screen time rather than longitudinal addictive use trajectories.

OBJECTIVES To identify trajectories of addictive use of social media, mobile phones, and video games and to examine their associations with suicidal behaviors and ideation and mental health outcomes among youths.

DESIGN, SETTING, AND PARTICIPANTS Cohort study analyzing data from baseline through year 4 follow-up in the Adolescent Brain Cognitive Development Study (2016-2022), with population-based samples from 21 US sites.

EXPOSURES Addictive use of social media, mobile phones, and video games using validated child-reported measures from year 2, year 3, and year 4 follow-up surveys.

MAIN OUTCOMES AND MEASURES Suicidal behaviors and ideation assessed using child- and parent-reported information via the Kiddie Schedule for Affective Disorders and Schizophrenia. Internalizing and externalizing symptoms were assessed using the parent-reported Child Behavior Checklist.

RESULTS The analytic sample (n = 4285) had a mean age of 10.0 (SD, 0.6) years; 47.9% were female; and 9.9% were Black, 19.4% Hispanic, and 58.7% White. Latent class linear mixed models identified 3 addictive use trajectories for social media and mobile phones and 2 for video games. Nearly one-third of participants had an increasing addictive use trajectory for social media or mobile phones beginning at age 11 years. In adjusted models, increasing addictive use trajectories were associated with higher risks of suicide-related outcomes than low addictive use trajectories (eg, increasing addictive use of social media had a risk ratio of 2.14 [95% CI, 1.61-2.85] for suicidal behaviors). High addictive u... ...ectories for all screen types were associated with suicide-related outcomes (eg, hi... ...ddictive use of social media had a risk ratio of 2.39 [95% CI, 1.66-3.43] for suicid... ...igh video game addictive use trajectory showed the largest relative diff... ...ptoms (T score difference, 2.03 [95% CI, 1.45-2.61]), and the ... ...use trajectory for externalizing symptoms (T score diff... compared with low addictive use trajectories. To... associated with outcomes.

CONCLUSIONS AND RELEVANCE High or inc... mobile phones, or video games were co... addictive screen use trajectories were... worse mental health.

Harm-

Author Affiliations: Department of Population Health Sciences, Weill ...ell Medicine/NewYork-Presbyterian, ...rk, New York (Xiao, Meng); ...blic Health, University of ...ley (Brown); ...Mailman School ...ew York, New York ...ments of Psychiatry ...g Irving Medical Center, ...niversity, New York, ...(Mann); Division of ...lar Imaging and ...ropathology, New York State ...ychiatric Institute, New York, New York (Mann).

Corresponding Author: Yunyu Xiao, PhD, Weill Cornell Medicine/NewYork-Presbyterian Department of Population Health Sciences, 575 Lexington Ave, FP650A, New York, NY 10022 (yux4008@med.cornell.edu).

JAMA. doi:10.1001/jama.2025.7829
Published online June 18, 2025.

JAMA | Original Investigation

Addictive Screen Use Trajectories and Suicidal Behaviors, Suicidal Ideation, and Mental Health in US Youths

Yunyu Xiao, PhD; Yuan Meng, PhD; Timothy T. Brown, PhD; Katherine M. Keyes, PhD; J. John Mann, MD

Editorial

Supplemental content

IMPORTANCE Increasing child and adolescent use of social media, video games, and mobile phones has raised concerns about potential links to youth mental health problems. Prior research has largely focused on total screen time rather than longitudinal addictive use trajectories.

OBJECTIVES To identify trajectories of addictive use of social media, mobile phones, and video games and to examine their associations with suicidal behaviors and ideation and mental health outcomes among youths.

DESIGN, SETTING, AND PARTICIPANTS Cohort study analyzing data from baseline through year 4 follow-up in the Adolescent Brain Cognitive Development Study (2016-2022), with population-based samples from 21 US sites.

EXPOSURES Addictive use of social media, mobile phones, and video games using validated child-reported measures from year 2, year 3, and year 4 follow-up surveys.

MAIN OUTCOMES AND MEASURES Suicidal behaviors and ideation assessed using child- and parent-reported information via the Kiddie Schedule for Affective Disorders and Schizophrenia. Internalizing and externalizing symptoms were assessed using the parent-reported Child Behavior Checklist.

RESULTS The analytic sample (n = 4285) had a mean age of 10.0 (SD, 0.6) years; 47.9% were female; and 9.9% were Black, 19.4% Hispanic, and 58.7% White. Latent class linear mixed models identified 3 addictive use trajectories for social media and mobile phones and 2 for video games. Nearly one-third of participants had an increasing addictive use trajectory for social media or mobile phones beginning at age 11 years. In adjusted models, increasing addictive use trajectories were associated with higher risks of suicide-related outcomes than low addictive use trajectories (eg, increasing addictive use of social media had a risk ratio of 2.14 [95% CI, 1.61-2.85] for suicidal behaviors). High addictive use trajectories for all screen types were associated with suicide-related outcomes (eg, high-peaking addictive use of social media had a risk ratio of 2.39 [95% CI, 1.66-3.43] for suicidal behaviors). The high video game addictive use trajectory showed the largest relative difference in internalizing symptoms (T score difference, 2.03 [95% CI, 1.45-2.61]), and the increasing social media addictive use trajectory for externalizing symptoms (T score difference, 1.05 [95% CI, 0.54-1.56]), compared with low addictive use trajectories. Total screen time at baseline was not associated with outcomes.

CONCLUSIONS AND RELEVANCE High or increasing trajectories of addictive use of social media, mobile phones, or video games were common in early adolescents. Both high and increasing addictive screen use trajectories were associated with suicidal behaviors and ideation and worse mental health.

Author Affiliations: Department of Population Health Sciences, Weill Cornell Medicine/NewYork-Presbyterian, New York, New York (Xiao, Meng); School of Public Health, University of California, Berkeley (Brown); Columbia University Mailman School of Public Health, New York, New York (Keyes); Departments of Psychiatry and Radiology, Irving Medical Center, Columbia University, New York, New York (Mann); Division of Molecular Imaging and Neuropathology, New York State Psychiatric Institute, New York, New York (Mann).

Corresponding Author: Yunyu Xiao, PhD, Weill Cornell Medicine/ NewYork-Presbyterian Department of Population Health Sciences, 575 Lexington Ave, FP6S0A, New York, NY 10022 (yux4008@med.cornell.edu).

JAMA. doi:10.1001/jama.2025.7829
Published online June 18, 2025.

The increasing use of social media, video games, mobile phones, and other screen-based activities among adolescents, combined with rising rates of suicidal behaviors and mental health problems in children and younger adolescents, has raised concerns,[1-4] including a US surgeon general warning label.[5] While most existing research has focused on total screen time,[6-11] emerging evidence suggests that addictive screen use may be a more salient risk factor for suicidality and mental health in youths.[12-15] Addictive use may vary by platform[16,17] and follow distinct developmental trajectories. However, addictive use trajectories among youths have not been well characterized, and how they may relate to suicide-related and mental health outcomes remains largely unknown.[18,19]

To address these gaps, this study used nationwide data from the Adolescent Brain Cognitive Development (ABCD) Study, a population-based, longitudinal cohort of children and adolescents, to (1) characterize longitudinal trajectories of addictive use of social media, mobile phones, and video games; (2) assess whether addictive use trajectories were associated with suicidal behaviors, suicidal ideation, and internalizing and externalizing symptoms over 4 years, controlling for baseline demographics and clinical characteristics; and (3) examine whether addictive use trajectories were associated with outcomes after adjusting for total screen time.

Key Points

Question Are addictive screen use trajectories associated with suicidal behaviors, suicidal ideation, and mental health outcomes in US youth?

Findings In this cohort study of 4285 US adolescents, 31.3% had increasing addictive use trajectories for social media and 24.6% for mobile phones over 4 years. High or increasing addictive use trajectories were associated with elevated risks of suicidal behaviors or ideation compared with low addictive use. Youths with high-peaking or increasing social media use or high video game use had more internalizing or externalizing symptoms.

Meaning Both high and increasing addictive screen use trajectories were associated with suicidal behaviors, suicidal ideation, and worse mental health in youths.

Methods

This study used the most recent available data from the ABCD Study (release 5.1),[20] a longitudinal cohort study of participants aged 9 to 10 years recruited from 21 US sites at baseline (n = 11 868) and followed up annually. Data collection spanned 2016 through January 2022, covering both the COVID-19 prepandemic and postpandemic years. Because the 4-year follow-up data release is ongoing, we used the available random subset (n = 4754). χ^2 Automatic interaction detection (CHAID) analysis comparing baseline characteristics of participants with and without year 4 follow-up data showed no selection bias (eAppendix 1 in Supplement 1).[21]

Our analytic sample included 4285 participants with complete data on addictive screen use from the year 2 to year 4 follow-up surveys and baseline demographics (Figure 1).[22-24] This study was approved by institutional review boards at each site, with central institutional review board approval from the University of California, San Diego. Parents or guardians provided written informed consent. This study follows the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline.

Addictive Screen Use (Years 2-4 Follow-Up)

Validated self-report questionnaires[25,26] were used to assess addictive uses for 3 platforms—social media, mobile phones, and video games—including a 6-item Social Media Addiction Questionnaire (SMAQ), 8-item Mobile Phone Involvement Questionnaire (MPIQ), and 6-item Video Game Addiction Questionnaire (VGAQ),[27,28] measuring compulsive use, difficulty disengaging, and distress when not using (see details in the Box and in eTable 1 in Supplement 1). Responses used Likert-type scales (1 ["never"] to 6 ["very often"] for SMAQ and VGAQ; 1 ["strongly disagree"] to 7 ["strongly agree"] for MPIQ). We calculated weighted addictive use scores using confirmatory factor analysis,[22] which were appropriate for this study because of their greater measurement precision and construct validity than mean scores (eTable 1 and eAppendix 2 in Supplement 1).[29] Higher scores indicate greater addictive use. All scales have high reliability (Cronbach α = 0.88 for each scale) (eTable 2 in Supplement 1).

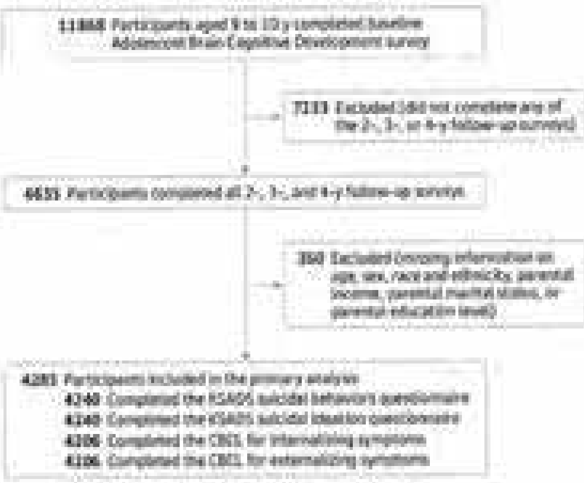
Total Screen Time (Baseline)

Because different screen activities can overlap, this analysis used self-reported questions assessing the average daily non-schoolwork-related screen time (separately for weekdays and weekends) (eTable 1 in Supplement 1). Prior ABCD studies showed positive correlations between self-reported and objectively measured screen use (r = 0.49; P < .001).[8]

Suicidal Behaviors and Suicidal Ideation (Year 4 Follow-Up)

Child and parent reports of suicidal behaviors and suicidal ideation over the prior year were assessed at year 4 follow-up using the Kiddie Schedule for Affective Disorders and Schizophrenia (KSADS),[30,31] covering a spectrum of suicide-related outcomes: (1) passive ideation; (2) nonspecific active suicidal ideation; (3) specific active suicidal ideation; (4) active ideation with intent; (5) active ideation with plan and intent; (6) preparatory actions for imminent suicidal behavior; (7) interrupted suicidal attempt; (8) aborted suicidal attempt; and (9) suicide attempt (eTable 3 in Supplement 1). Consistent with prior literature,[32,33] suicidal ideation was classified as present if any of items 1 to 5 were endorsed, and suicidal behaviors were classified as present if any of items 6 to 9 were endorsed, by either the youth or caregiver. The KSADS exhibits strong validity and reliability for this population.[30]

Figure 1. Participant Flow in a Study of Addictive Screen Use Trajectories and Suicidal Behaviors and Ideation and Mental Health in Youths



CBCL indicates Child Behavior Checklist; KSADS, Kiddie Schedule for Affective Disorders and Schizophrenia.

Mental Health Outcomes (Year 4 Follow-Up)

Analyses included current parent-reported internalizing (eg, anxiety, depression) and externalizing (eg, aggression, rule-breaking) symptoms using T scores derived from the Child Behavior Checklist (CBCL).[29] T scores of 65 or greater are considered to indicate clinically elevated symptoms.[34]

Covariates (Baseline)

Models were adjusted for child age, sex, race and ethnicity, parental income, education, and marital status as reported in the baseline ABCD parent demographics survey. Race and ethnicity were based on caregiver-reported predefined categories, including non-Hispanic Asian, non-Hispanic Black, Hispanic (any race), non-Hispanic White, and multiracial and/or other racial and ethnic groups, collected as social constructs to investigate the differential impact of structural disadvantages (eTable 3 in Supplement 1).[35,36] Models were also adjusted for baseline clinical characteristics (ie, suicidal behaviors, suicidal ideation, and internalizing and externalizing symptoms).

Statistical Analysis

Latent class linear mixed models[37] were used to identify addictive use trajectories based on age and quadratic age terms (eAppendix 3 in Supplement 1). In the addictive use questionnaires, missing data resulting from skip patterns based on previous use or nonuse questions were replaced with "1 = never/strongly disagree" for participants who reported having no social media accounts or mobile phones or who did not play video games. The optimal group-based trajectory model was selected based on (1) the lowest bayesian information criterion; (2) greater than 70% average probability of participants being correctly classified into their respective trajectories; (3) greater than 5.0 odds of correct classification; and (4) greater than 5% minimum trajectory sample sizes out of the total sample. Each addictive use trajectory represents children who shared similar addictive use levels over time.

Subsequently, these addictive use trajectories were treated as categorical variables in outcome models. For categorical outcomes (suicidal behaviors and suicidal ideation), Poisson regression models were used to estimate risk ratios (RRs), with 95% CIs calculated using robust standard errors. Poisson models are appropriate in this context.[38-40] For continuous outcomes (internalizing and externalizing symptoms), generalized linear models were used to estimate mean differences, with 95% CIs calculated using ordinary standard errors. E-values were computed to evaluate sensitivity to unmeasured confounding.[41] Total screen time was added as a covariate to examine whether it explained the magnitude or direction of the associations between addictive use trajectories and outcomes. In the sensitivity analysis, total screen time was also tested for independent associations with the outcomes.

Significance was set at a 2-sided $P < .05$, corrected for false discovery rate (FDR) using the Benjamini-Hochberg method to adjust for multiple testing within each group.[42] All analyses were conducted in R version 4.3.1 (R Foundation).

Box. Sample Items From Addictive Use Scales and Baseline Screen Time Measures[a]

Addictive Use (Social Media/Mobile Phone/Video Games)

- I feel the need to use social media apps more and more (1 [never] to 6 [very often]).
- The thought of being without my phone makes me feel distressed (1 [strongly disagree] to 7 [strongly agree]).
- I play video games so I can forget about my problems (1 [never] to 6 [very often]).

Total Screen Time (Weekday/Weekend)

Total typical weekend and weekday screen times on streaming movies or television shows, single-player games, multiplayer games, texting, social media, and video chatting (0-24 hours).

[a] eTable 1 in Supplement 1 is a full table of addictive use and screen time measures.

Results

Among 4285 participants (baseline mean age, 10.0 [SD, 0.6] years; 47.9% female), the sample included 96 (2.2%) Asian, 426 (9.9%) Black, 830 (19.4%) Hispanic, and 2515 (58.7%) White individuals, as well as 418 (9.8%) individuals identifying as multiracial and/or other races (Table).

Addictive Use Trajectories

The optimal trajectory models were selected based on multiple fit criteria (eTable 4 and eAppendix 4 in Supplement 1).

For social media, 3 addictive use trajectories emerged (Figure 2): high-peaking (n = 410 [9.6%]), increasing (n = 1342 [31.3%]), and low (n = 2533 [59.1%]). At baseline, high-peaking and increasing trajectories had similar levels of

Table. Baseline Characteristics and Year 4 Follow-Up Suicidal Behaviors, Suicidal Ideation, and Mental Health Outcomes

Characteristics	No. (%) (n = 4285)
Age, mean (SD), y	10.0 (0.6)
Sex	
Female	2051 (47.9)
Male	2234 (52.1)
Race and ethnicity	
Asian	96 (2.2)
Black	426 (9.9)
Hispanic	830 (19.4)
White	2515 (58.7)
Multiracial and/or other[a]	418 (9.8)
Annual household income, $	
<75 000	1722 (40.2)
≥75 000	2563 (59.8)
Parental marital status	
Married	3138 (73.2)
Living with partner	197 (4.6)
Single	950 (22.2)
Parental education	
Less than bachelor's degree	1474 (34.4)
Bachelor's or higher	2811 (65.6)
Suicidal behaviors (year 4 follow-up)[b]	
No	4036 (94.9)
Yes	218 (5.1)
Suicidal ideation (year 4 follow-up)[c]	
No	3494 (82.1)
Yes	760 (17.9)
Child Behavior Checklist T score, mean (SD)[d]	
Internalizing symptoms	47.5 (10.8)
Externalizing symptoms	43.4 (9.3)

[a] Primary caregivers were allowed to choose multiple race subgroups for children; the "other" category indicates that no specific race or ethnicity group was identified.

[b] Suicidal behaviors were determined if any of the questions for the following Kiddie Schedule for Affective Disorders and Schizophrenia (KSADS) items had a "yes" response by the child or the caregiver: (1) preparatory actions for imminent suicidal behavior, (2) interrupted suicidal attempt, (3) aborted suicidal attempt, and (4) suicide attempt.

[c] Suicidal ideation was determined if any of the questions for the following KSADS items had a "yes" response by the child or the caregiver: (1) passive ideation, (2) nonspecific active suicidal ideation, (3) specific active suicidal ideation, (4) active ideation with intent, and (5) active ideation with plan and intent.

[d] Child Behavior Checklist T scores for internalizing range from 33 to 87 and for externalizing range from 33 to 82. T scores are standard scores derived from raw scores. Higher T scores reflect more severe internalizing and externalizing symptoms; T scores of 65 or greater indicate clinically meaningful internalizing and externalizing concerns.

social media addictive use, providing no clear indication of their subsequent divergence. By age 14 years, the increasing social media addictive use trajectory reached levels comparable with the high-peaking addictive use trajectory and continued to rise further.

Mobile phone addictive use also followed 3 trajectories: high (n = 2109 [49.2%]), increasing (n = 1052 [24.6%]), and low (n = 1124 [26.2%]). The low and increasing addictive use trajectories began with almost the same baseline levels but diverged in their subsequent trajectories. The increasing mobile phone addictive use trajectory showed a steady increase in its addictive use level in the following 4 years, reaching levels comparable with the high addictive use trajectory by age 15 years.

For video games, 2 trajectories were identified: high addictive use (n = 1761 [41.1%]) and low addictive use (n = 2524 [58.9%]).

Trajectory Differences in Baseline Demographics and Clinical Characteristics

The high addictive social media use trajectory included a higher proportion of females than the low addictive use trajectory (51.0% vs 42.8%; absolute difference, 8.18%; 95% CI, 3.07%-13.36%) (eTable 5 in Supplement 1). In contrast, youths in high addictive video game use trajectories were more likely to be male than those in low addictive use trajectories (70.1% vs 39.6%; absolute difference, 30.55%; 95% CI, 27.64%-33.40%).

Youths in high addictive use trajectories were more likely to be Black (absolute differences, 3.08%-7.91%) or Hispanic (absolute differences, 7.12%-10.03%) compared with those in low addictive use trajectories.

High addictive use trajectories also had higher proportions of youths from households with annual incomes below $75 000, unmarried parents, and parents with less than a bachelor's degree education (absolute differences across these indicators ranged from 1.55% to 18.95%) compared with low addictive use trajectories.

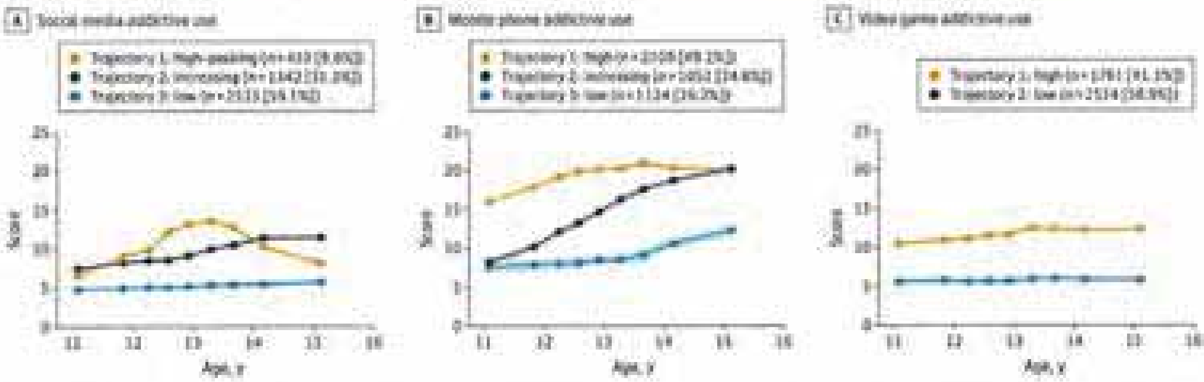
Youths in high addictive social media use trajectories had the largest differences in baseline suicidal behaviors (absolute difference, 1.67%; 95% CI, 0.06%-3.50%) and baseline externalizing symptom scores (absolute T score mean difference, 1.79; 95% CI, 0.75-2.82) compared with those in low addictive use trajectories. The largest differences in baseline suicidal ideation (absolute difference, 6.79%; 95% CI, 4.58%-8.91%) and baseline internalizing symptom scores (absolute T score mean difference, 1.80; 95% CI, 1.19-2.44) were observed between the groups in high and low addictive use trajectories for video games.

Associations of Trajectories of Addiction Severity With Suicidal Behaviors, Suicidal Ideation, and Mental Health

Among 4285 participants, 218 (5.1%) reported suicidal behaviors and 760 (17.9%) reported suicidal ideation at year 4 follow-up. Mean year 4 CBCL internalizing and externalizing T scores were 47.5 (SD, 10.8) and 43.4 (SD, 9.3), respectively (Table).

For social media addictive use, adjusted models (Figure 3) showed that both high-peaking and increasing addictive use trajectories were associated with higher risk of suicidal behaviors (high-peaking: RR, 2.39; 95% CI, 1.66-3.43; FDR-adjusted $P < .001$; increasing: RR, 2.14; 95% CI, 1.61-2.85; FDR-adjusted $P < .001$) and elevated risk of suicidal ideation (high-peaking: RR, 1.51; 95% CI, 1.25-1.83; FDR-adjusted $P < .001$; increasing: RR, 1.46; 95% CI, 1.28-1.67;

Figure 2. Addictive Use Trajectories of Social Media, Mobile Phones, and Video Games



Latent class linear mixed models were used to identify distinct trajectories for each type of addictive use based on repeated measures of self-reported use of social media, mobile phones, and video games from ages 11 to 15 years. Each trajectory represents a group of children with similar temporal patterns of addictive use. Models were fit separately for each screen type and regressed on age and quadratic age terms. Model selection was based on the lowest bayesian information criterion, an average posterior probability of assignment greater than 70%, an odds of correct classification greater than 5.0, and a minimum group size of 5% (eTable 4 in Supplement 1). Addictive use scores were derived from confirmatory factor analysis (eTable 2 and eAppendix 2 in Supplement 1) and ranged as follows: social media, 4.5-26.8, mobile phone, 5.6-39.5, and video games, 4.5-26.9. Shaded areas represent 95% CIs. Data points along each trajectory line represent model-estimated mean scores at specific ages based on the latent class linear mixed models. Age values reflect quantiles of the observed age distribution.

FDR-adjusted $P = .001$) compared with the low addictive use trajectory. Internalizing symptom T scores were higher in the increasing addictive use trajectory (mean difference, 1.27; 95% CI, 0.66-1.88; FDR-adjusted $P < .001$), while externalizing symptom T scores were higher in both high-peaking (mean difference, 1.25; 95% CI, 0.45-2.04; FDR-adjusted $P = .004$) and increasing (mean difference, 1.05; 95% CI, 0.54-1.56; FDR-adjusted $P < .001$) addictive use trajectories compared with the low addictive use trajectory, all having small effect sizes (Cohen $d < .2$).

For mobile phone use, the high addictive use trajectory was associated with higher risks of suicidal behaviors (RR, 2.17; 95% CI, 1.48-3.19; FDR-adjusted $P < .001$) and suicidal ideation (RR, 1.50; 95% CI, 1.27-1.78; FDR-adjusted $P < .001$) compared with the low addictive use trajectory. The increasing addictive use trajectory was modestly associated with a greater relative risk of suicidal ideation (RR, 1.22; 95% CI, 1.01-1.48; FDR-adjusted $P < .001$) but not with other mental health outcomes.

For video game addictive use, the high addictive use trajectory was associated with higher risk of suicidal behaviors (RR, 1.54; 95% CI, 1.18-2.03; FDR-adjusted $P = .004$) and suicidal ideation (RR, 1.53; 95% CI, 1.35-1.75; FDR-adjusted $P < .001$), as well as higher internalizing symptom T scores (mean difference, 2.03; 95% CI, 1.45-2.61; FDR-adjusted $P < .001$) and externalizing symptom T scores (mean difference, 0.94; 95% CI, 0.45-1.43; FDR-adjusted $P < .001$) compared with the low addictive use trajectory.

Baseline total screen time alone was not associated with suicidal behaviors, suicidal ideation, or internalizing or externalizing symptom associations (Figure 4). Additionally, when models were adjusted for addictive use trajectories, baseline screen time remained not independently associated with these outcomes (eFigures 1-3 in Supplement 1). E-values indicated moderate to strong robustness to potential unmeasured confounding (range, 1.10-4.21).

Discussion

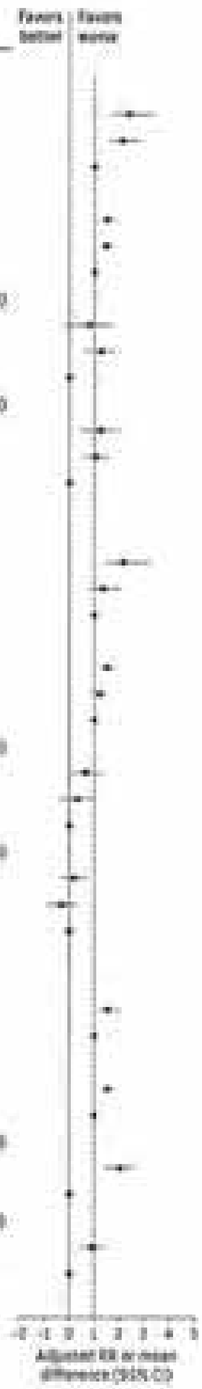
This study identified distinct trajectories of addictive use of social media, mobile phones, and video games from childhood to early adolescence and found links to suicidal behaviors, suicidal ideation, and worse mental health outcomes. High or increasing addictive use trajectories were common. Almost 1 in 2 youths had a high addictive use trajectory for mobile phones, and more than 40% had a high addictive use trajectory for video games. Many others had increasing addictive use over the 4-year observation period that ended with high addictive use; almost 1 in 3 had this trajectory for social media and 1 in 4 for mobile phones.

For social media and mobile phones, both the high and increasing addictive use trajectories were associated with 2 to 3 times greater risks of suicidal behaviors and suicidal ideation compared with the low addictive use trajectory. High-peaking and increasing addictive use trajectories of social media were also associated with higher internalizing and externalizing symptom scores compared with the low addictive use trajectory. For video games, the high addictive use trajectory was associated with greater risks of suicidal behaviors, suicidal ideation, and higher internalizing symptoms scores compared with the low addictive use trajectory.

To our knowledge, this is the first study to characterize longitudinal addictive use trajectories for social media, mobile phones, and video games among children and early adolescents and to assess their prospective associations with suicide-related and mental health outcomes. Specific strengths include

Figure 3. Associations of Addictive Use Trajectories With Year 4 Follow-Up Suicidal Behaviors, Suicidal Ideation, and Mental Health Outcomes

Outcomes, by trajectory[a]	No./total No. (%) With outcome	No./total No. (%) Without outcome	Unadjusted RR or mean difference (95% CI)[b]	Adjusted RR or mean difference (95% CI)[b]
Social media addictive use				
Suicidal behaviors			RR (95% CI)	RR (95% CI)
High-peaking	36/407 (8.8)	371/407 (91.2)	3.00 (2.05 to 4.38)	2.39 (1.66 to 3.43)
Increasing	105/1326 (7.9)	1221/1326 (92.1)	2.61 (1.96 to 3.48)	2.14 (1.61 to 2.85)
Low	76/2507 (3.0)	2431/2507 (97.0)	1 [Reference]	1 [Reference]
Suicidal ideation			RR (95% CI)	RR (95% CI)
High-peaking	94/407 (23.1)	313/407 (76.9)	1.64 (1.34 to 2.00)	1.51 (1.25 to 1.83)
Increasing	305/1326 (23.0)	1021/1326 (77.0)	1.62 (1.41 to 1.86)	1.46 (1.28 to 1.67)
Low	357/2507 (14.2)	2150/2507 (85.8)	1 [Reference]	1 [Reference]
CBCL internalizing			Mean difference (95% CI)	Mean difference (95% CI)
High-peaking	36/403 (8.9)	367/403 (91.1)	0.93 (−0.20 to 2.07)	0.78 (−0.17 to 1.73)
Increasing	138/1315 (10.5)	1177/1315 (89.5)	1.33 (0.61 to 2.05)	1.27 (0.68 to 1.86)
Low	165/2488 (6.6)	2323/2488 (93.4)	[Reference]	[Reference]
CBCL externalizing			Mean difference (95% CI)	Mean difference (95% CI)
High-peaking	25/403 (6.2)	378/403 (93.8)	2.03 (1.06 to 3.00)	1.25 (0.45 to 2.04)
Increasing	38/1315 (2.9)	1277/1315 (97.1)	1.18 (0.57 to 1.80)	1.05 (0.54 to 1.56)
Low	45/2488 (1.8)	2443/2488 (98.2)	[Reference]	[Reference]
Mobile phone addictive use				
Suicidal behaviors			RR (95% CI)	RR (95% CI)
High	146/2085 (7.0)	1939/2085 (93.0)	2.52 (1.72 to 3.69)	2.17 (1.48 to 3.19)
Increasing	41/1040 (3.9)	999/1040 (96.1)	1.43 (0.89 to 2.24)	1.35 (0.86 to 2.12)
Low	30/1115 (2.7)	1085/1115 (97.3)	1 [Reference]	1 [Reference]
Suicidal ideation			RR (95% CI)	RR (95% CI)
High	435/2085 (20.9)	1650/2085 (79.1)	1.57 (1.33 to 1.86)	1.50 (1.27 to 1.78)
Increasing	173/1040 (16.6)	867/1040 (83.4)	1.26 (1.03 to 1.54)	1.22 (1.01 to 1.48)
Low	148/1115 (13.3)	967/1115 (86.7)	1 [Reference]	1 [Reference]
CBCL internalizing			Mean difference (95% CI)	Mean difference (95% CI)
High	180/2067 (8.7)	1887/2067 (91.3)	0.58 (−0.20 to 1.37)	0.66 (0.00 to 1.33)
Increasing	79/1030 (7.7)	951/1030 (92.3)	−0.04 (−0.93 to 0.88)	0.32 (−0.44 to 1.08)
Low	80/1109 (7.2)	1029/1109 (92.8)	[Reference]	[Reference]
CBCL externalizing			Mean difference (95% CI)	Mean difference (95% CI)
High	57/2067 (2.8)	2010/2067 (97.2)	0.43 (−0.25 to 1.10)	0.20 (−0.36 to 0.76)
Increasing	20/1030 (1.9)	1010/1030 (98.1)	−0.66 (−1.44 to 0.13)	−0.30 (−0.94 to 0.33)
Low	31/1109 (2.8)	1078/1109 (97.2)	[Reference]	[Reference]
Video game addictive use				
Suicidal behaviors			RR (95% CI)	RR (95% CI)
High	106/1749 (6.1)	1643/1749 (93.9)	1.38 (1.06 to 1.78)	1.54 (1.18 to 2.03)
Low	111/2491 (4.5)	2380/2491 (95.5)	1 [Reference]	1 [Reference]
Suicidal ideation			RR (95% CI)	RR (95% CI)
High	366/1749 (20.9)	1383/1749 (79.1)	1.34 (1.18 to 1.52)	1.53 (1.35 to 1.75)
Low	390/2491 (15.7)	2101/2491 (84.3)	1 [Reference]	1 [Reference]
CBCL internalizing			Mean difference (95% CI)	Mean difference (95% CI)
High	178/1728 (10.3)	1550/1728 (89.7)	1.83 (1.37 to 2.49)	2.03 (1.45 to 2.61)
Low	161/2478 (6.5)	2317/2478 (93.5)	[Reference]	[Reference]
CBCL externalizing			Mean difference (95% CI)	Mean difference (95% CI)
High	58/1728 (3.4)	1670/1728 (96.6)	1.71 (1.15 to 2.28)	0.94 (0.45 to 1.43)
Low	50/2478 (2.0)	2428/2478 (98.0)	[Reference]	[Reference]



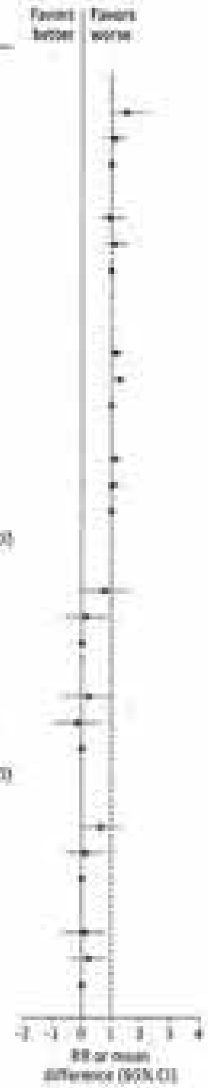
Dashed vertical line at x = 1 represents the reference for risk ratios (RRs). Solid vertical line at x = 0 represents the reference for mean differences.

[a] See descriptions of suicidal behaviors, suicidal ideation, and Child Behavior Checklist (CBCL) internalizing and externalizing scores in footnotes b-d of the Table. Participants with CBCL internalizing and externalizing T scores ≥ 65 are shown in the "With outcome" column as these scores are considered to indicate clinically elevated symptoms. Exposure categories were dummy coded (low, increasing, high), with the low-use trajectory as the reference group.

[b] For categorical outcomes (suicidal behaviors and suicidal ideation), Poisson regression was used to estimate RRs and 95% CIs using robust standard errors. For continuous outcomes (internalizing and externalizing symptoms), generalized linear models were used to estimate mean differences with 95% CIs using ordinary standard errors. Unadjusted models included only the addictive use trajectories. Adjusted models also controlled for baseline age; sex; race and ethnicity; parental education, income, and marital status; baseline suicidal ideation and behaviors; and baseline internalizing and externalizing symptoms.

Figure 4. Associations of Baseline Screen Time With Year 4 Follow-Up Suicidal Behaviors, Suicidal Ideation, and Mental Health Outcomes

Outcomes, by baseline screen time[a]	No./total No. (%) With outcome	Without outcome	RR or mean difference (95% CI)[b]
Suicidal behaviors			RR (95% CI)
Weekday screen time			
High	78/1138 (6.9)	1060/1138 (93.1)	1.53 (1.04 to 2.26)
Moderate	54/1133 (4.8)	1079/1133 (95.2)	1.09 (0.78 to 1.53)
Low	85/1964 (4.3)	1879/1964 (95.7)	1 [Reference]
Weekend screen time			
High	99/1732 (5.7)	1633/1732 (94.3)	0.93 (0.64 to 1.41)
Moderate	65/1265 (5.1)	1200/1265 (94.9)	1.11 (0.79 to 1.58)
Low	53/1238 (4.3)	1185/1238 (95.7)	1 [Reference]
Suicidal ideation			RR (95% CI)
Weekday screen time			
High	212/1138 (18.6)	926/1138 (81.4)	1.16 (0.95 to 1.40)
Moderate	227/1133 (20.0)	906/1133 (80.0)	1.26 (1.08 to 1.48)
Low	315/1964 (16.0)	1649/1964 (84.0)	1 [Reference]
Weekend screen time			
High	332/1732 (19.2)	1400/1732 (80.8)	1.10 (0.91 to 1.32)
Moderate	225/1265 (17.8)	1040/1265 (82.2)	1.04 (0.87 to 1.23)
Low	197/1238 (15.9)	1041/1238 (84.1)	1 [Reference]
CBCL internalizing			Mean difference (95% CI)
Weekday screen time			
High	106/1130 (9.4)	1024/1130 (90.6)	0.77 (−0.08 to 1.62)
Moderate	84/1123 (7.5)	1039/1123 (92.5)	0.15 (−0.38 to 0.68)
Low	149/1948 (7.6)	1799/1948 (92.4)	[Reference]
Weekend screen time			
High	157/1718 (9.1)	1561/1718 (90.9)	0.22 (−0.62 to 1.06)
Moderate	84/1261 (6.7)	1177/1261 (93.3)	−0.14 (−0.88 to 0.61)
Low	98/1222 (8.0)	1124/1222 (92.0)	[Reference]
CBCL externalizing			Mean difference (95% CI)
Weekday screen time			
High	45/1130 (4.0)	1085/1130 (96.0)	0.65 (−0.08 to 1.37)
Moderate	24/1123 (2.1)	1099/1123 (97.9)	0.12 (−0.48 to 0.74)
Low	39/1948 (2.0)	1909/1948 (98.0)	[Reference]
Weekend screen time			
High	54/1718 (3.1)	1664/1718 (96.9)	0.11 (−0.60 to 0.81)
Moderate	28/1261 (2.2)	1233/1261 (97.8)	0.22 (−0.41 to 0.84)
Low	26/1222 (2.1)	1196/1222 (97.9)	[Reference]



Risk estimates from models examining associations between baseline screen time (weekday and weekend) and year 4 outcomes: suicidal behaviors, suicidal ideation, and Child Behavior Checklist (CBCL) internalizing and externalizing symptoms, controlling for demographics, suicidal behaviors, suicidal ideation, and CBCL internalizing and externalizing symptom T scores at baseline. Dashed vertical line at x = 1 represents the reference for risk ratios (RRs). Solid vertical line at x = 0 represents the reference for mean differences.

[a] See descriptions of suicidal behaviors, suicidal ideation, and CBCL internalizing and externalizing T scores in footnotes b-d of the Table. Participants with CBCL internalizing and externalizing T scores ≥65 are shown in the "With outcome" column as these scores are considered to indicate clinically elevated symptoms. Baseline screen time was classified as low (<2 h/d),[47] moderate (≥2 to <4 h/d), or high (≥4 h/d).[54] Cutoffs were selected based on existing literature that has linked moderate and high levels of screen time to elevated risks of depressive symptoms, anxiety, and behavioral problems in children and adolescents. Exposure categories were dummy coded (low, high, medium), with low screen time as the reference group.

[b] For categorical outcomes (suicidal behaviors, suicidal ideation), Poisson regression was used to estimate RRs and 95% CIs using robust standard errors. For continuous outcomes (internalizing and externalizing symptoms), generalized linear models were used to estimate mean differences with 95% CIs using ordinary standard errors.

the use of a large, population-based longitudinal sample and comprehensive, platform-specific assessment of addictive use trajectories. Previous studies, mostly cross-sectional and measuring only total screen time, have reported associations between more screen time and poorer mental health.[4,41,45,48] The current study's findings align with prior studies observing associations between addictive screen use and psychiatric symptoms at single time points.[47,49] This study adds substantially to existing knowledge by examining longitudinal trajectories and their associations with long-term outcomes.

For both social media and mobile phones, addictive use trajectories followed 3 different patterns, and a substantial proportion of youths had addictive use trajectories that increased over the 4 years of observation, starting at age 10

years. These increasing addictive use patterns, which would not have been predicted based on baseline assessments alone, were associated with elevated risks of suicidal behaviors and ideation. This underscores the potential importance of repeated assessment of addictive use of social media and mobile phones among children entering adolescence. In contrast, video game addictive use followed 2 trajectories, high and low, which were stable over time, potentially allowing earlier identification of risk without repeated assessment.

One key finding was that total screen time was not associated with suicide-related or mental health outcomes, nor did it alter the strength or direction of associations between addictive use trajectories and these outcomes. This underscores the importance of treating time spent and addictive use as separate constructs when examining associations with suicide-related and mental health outcomes.[19]

These findings suggest that focusing future research or interventions on addictive screen use might hold more promise than focusing on total screen time, which may unnecessarily involve low-risk youths. Future studies could evaluate whether monitoring addictive screen use is useful to identify higher-risk youths in clinical practice. Future research could also evaluate interventions that address the addictive aspect of screen use and prevention approaches targeting higher-risk subgroups of children and adolescents.[49,50]

Limitations

There are limitations. First, the observational nature of this study precludes establishing that addictive use trajectories cause the outcomes studied, although the longitudinal design mitigates concerns about reverse causality. Second, reliance on self-reported data may introduce recall and social desirability biases.[20,51] This analysis used weighted confirmatory factor analysis scores for quantifying addictive use, confirming their construct validity, and personal estimates of screen use. Still, future studies should consider incorporating objective measures, such as passive digital monitoring. Third, parent-reported CBCL measures may underestimate mental health conditions. Fourth, the COVID-19 pandemic may have influenced screen time,[10] but sensitivity analyses demonstrated consistent findings (eAppendix 5 in Supplement 1). Fifth, the ABCD Study did not assess multitasking across screen platforms, so it is not possible to tell how measurement of multitasking would have affected these findings. Sixth, not all of the participants in the ABCD Study had year 4 follow-up data available at the time of this study; future analyses should seek to replicate results when these data are available. Finally, these analyses did not include psychosocial and behavioral factors such as bullying,[52] adverse childhood experiences,[53,54] parental monitoring,[55,56] sleep disturbances,[57] stress,[58] social isolation,[48] and social determinants of health (eg, neighborhood and school contexts).[59,60] Future studies should examine potential interactions and mediating relationships among these factors, addictive use trajectories, and mental health outcomes.

Conclusions

High or increasing trajectories of addictive use of social media, mobile phones, or video games were common in early adolescence and were associated with suicide-related and mental health outcomes. Addictive screen use trajectories warrant further study regarding potential use for clinical evaluation of risk and for the design and testing of interventions to improve youth mental health.

ARTICLE INFORMATION

Accepted for Publication: April 30, 2025.

Published Online: June 18, 2025. doi:10.1001/jama.2025.7829

Author Contributions: Drs Xiao and Meng had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Drs Xiao and Meng are co-first authors.
Concept and design: Xiao, Meng, Mann.
Acquisition, analysis, or interpretation of data: All authors.
Drafting of the manuscript: Xiao, Meng.
Critical review of the manuscript for important intellectual content: All authors.
Statistical analysis: Xiao, Meng, Brown.
Obtained funding: Xiao.
Administrative, technical, or material support: Xiao, Meng.
Supervision: Xiao, Mann.

Conflict of Interest Disclosures: Dr Mann reports receipt of royalties for commercial use of the Columbia-Suicide Severity Rating Scale from the Research Foundation for Mental Hygiene and the Columbia Pathways App from Columbia University. No other disclosures were reported.

Funding/Support: This study was supported by funding from the National Institute of Mental Health (grant RF1MH134649 to Dr Xiao), the American Foundation for Suicide Prevention (grant YIG-2-133-22 to Dr Xiao), and Google (to Dr Xiao).

Role of the Funder/Sponsor: The study supporters had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Data Sharing Statement: See Supplement 2.

REFERENCES

1. Orben A, Blakemore SJ. How social media affects teen mental health: a missing link. *Nature*. 2023;614(7948):410-412. doi:10.1038/d41586-023-00402-9

2. Abbasi J. Surgeon general sounds the alarm on social media use and youth mental health crisis. *JAMA*. 2023;330(1):11-12. doi:10.1001/jama.2023.10262

3. Nagata JM, Smith N, Alsamman S, et al. Association of physical activity and screen time with body mass index among US adolescents. *JAMA Netw Open*. 2023;6(2):e2255466. doi:10.1001/jamanetworkopen.2022.55466

4. Eirich R, McArthur BA, Anhorn C, McGuinness C, Christakis DA, Madigan S. Association of screen time with internalizing and externalizing behavior problems in children 12 years or younger: a systematic review and meta-analysis. *JAMA Psychiatry*. 2022;79(5):393-405. doi:10.1001/jamapsychiatry.2022.0155

5. Office of the Surgeon General. *Social Media and Youth Mental Health: The US Surgeon General's Advisory*. Published 2023. Accessed August 22, 2024. https://www.hhs.gov/sites/default/files/sg-youth-mental-health-social-media-advisory.pdf

6. Wade NE, Ortigara JM, Sullivan RM, et al; ABCD Novel Technologies Workgroup. Passive sensing of preteens' smartphone use: an Adolescent Brain Cognitive Development (ABCD) cohort substudy. *JMIR Ment Health*. 2021;8(10):e29426. doi:10.2196/29426

7. Odgers CL, Schueller SM, Ito M. Screen time, social media use, and adolescent development. *Annu Rev Dev Psychol*. 2020;2:485-502. doi:10.1146/annurev-devpsych-121318-084815

8. Stiglic N, Viner RM. Effects of screentime on the health and well-being of children and adolescents:

a systematic review of reviews. BMJ Open. 2019;9(1):e023191. doi:10.1136/bmjopen-2018-023191

9. Zhang Y, Choi KW, Delaney SW, Ge T, Pingault JB, Tiemeier H. Shared genetic risk in the association of screen time with psychiatric problems in children. JAMA Netw Open. 2023;6(11):e2341502. doi:10.1001/jamanetworkopen.2023.41502

10. Nagata JM, Al-Shoaibi AA, Leong AW, et al. Screen time and mental health: a prospective analysis of the Adolescent Brain Cognitive Development (ABCD) Study. BMC Public Health. 2024;24(1):2686. doi:10.1186/s12889-024-20102-x

11. Schmidt-Persson J, Rasmussen MG, Sørensen SO, et al. Screen media use and mental health of children and adolescents: a secondary analysis of a randomized clinical trial. JAMA Netw Open. 2024;7(7):e2419881. doi:10.1001/jamanetworkopen.2024.19881

12. Montag C, Demetrovics Z, Elhai JD, et al. Problematic social media use in childhood and adolescence. Addict Behav. 2024;153:107980. doi:10.1016/j.addbeh.2024.107980

13. Allcott H, Braghieri L, Eichmeyer S, Gentzkow M. The welfare effects of social media. Am Econ Rev. 2020;110(3):629-676. doi:10.1257/aer.20190658

14. Boer M, Stevens GW, Finkenauer C, van den Eijnden RJ. The course of problematic social media use in young adolescents: a latent class growth analysis. Child Dev. 2022;93(2):e168-e187. doi:10.1111/cdev.13712

15. Christakis DA, Hale L. Toward defining problematic media usage patterns in adolescents. JAMA. Published online April 28, 2025. doi:10.1001/jama.2025.6113

16. Course-Choi J, Hammond L. Social media use and adolescent well-being: a narrative review of longitudinal studies. Cyberpsychol Behav Soc Netw. 2021;24(4):223-236. doi:10.1089/cyber.2020.0020

17. Vannucci A, Johnson A, Villanti AC, Evans WD, Turner M. Patterns of social media use and their relationship to health risks among young adults. J Adolesc Health. 2019;64(2):158-164. doi:10.1016/j.jadohealth.2018.06.025

18. Nagata JM, Chu J, Ganson KT, et al. Contemporary screen time modalities and disruptive behavior disorders in children: a prospective cohort study. J Child Psychol Psychiatry. 2023;64(1):125-135. doi:10.1111/jcpp.13673

19. Chen YY, Chen F, Wu KC, Lu TH, Chi YC, Yip PS. Dynamic reciprocal relationships between traditional media reports, social media postings, and youth suicide in Taiwan between 2012 and 2021. SSM Popul Health. 2023;24:101543. doi:10.1016/j.ssmph.2023.101543

20. Saragosa-Harris NM, Chaku N, MacSweeney N, et al. A practical guide for researchers and reviewers using the ABCD Study and other large longitudinal datasets. Dev Cogn Neurosci. 2022;55:101115. doi:10.1016/j.dcn.2022.101115

21. Metzner C, Schilling A, Traxdorf M, et al. Classification at the accuracy limit: facing the problem of data ambiguity. Sci Rep. 2022;12(1):22121. doi:10.1038/s41598-022-26498-z

22. Bagot KS, Tomko RL, Marshall AT, et al. Youth screen use in the ABCD Study. Dev Cogn Neurosci. 2022;57:101150. doi:10.1016/j.dcn.2022.101150

23. Barch DM, Albaugh MD, Avenevoli S, et al. Demographic, physical and mental health assessments in the adolescent brain and cognitive development study: Rationale and description. Dev Cogn Neurosci. 2018;32:55-66. doi:10.1016/j.dcn.2017.10.010

24. Barch DM, Albaugh MD, Baskin-Sommers A, et al. Demographic and mental health assessments in the adolescent brain and cognitive development study: updates and age-related trajectories. Dev Cogn Neurosci. 2021;52:101031. doi:10.1016/j.dcn.2021.101031

25. Andreassen CS. Online social network site addiction: a comprehensive review. Curr Addict Rep. 2015;2(2):175-184. doi:10.1007/s40429-015-0056-9

26. Nagata JM, Lee CM, Yang J, et al. Associations between sexual orientation and early adolescent screen use: findings from the Adolescent Brain Cognitive Development (ABCD) Study. Ann Epidemiol. 2023;82:54-58. doi:10.1016/j.annepidem.2023.03.004

27. Walsh SP, White KM, Young RM. Needing to connect: the effect of self and others on young people's involvement with their mobile phones. Aust J Psychol. 2010;62(4):194-203. doi:10.1080/00049530903567229

28. Andreassen CS, Torsheim T, Brunborg GS, Pallesen S. Development of a Facebook addiction scale. Psychol Rep. 2012;110(2):501-517. doi:10.2466/02.09.18.PR0.110.2.501-517

29. Achenbach TM. The Achenbach System of Empirically Based Assessment (ASEBA): Development, Findings, Theory, and Applications. Published January 22, 2019. Accessed August 23, 2024. https://aseba.org/the-achenbach-system-of-empirically-based-assessment/

30. Townsend L, Kobak K, Kearney C, et al. Development of three web-based computerized versions of the Kiddie Schedule for Affective Disorders and Schizophrenia child psychiatric diagnostic interview: preliminary validity data. J Am Acad Child Adolesc Psychiatry. 2020;59(2):309-325. doi:10.1016/j.jaac.2019.05.009

31. Kaufman J, Kobak K, Birmaher B, de Lacy N. KSADS-COMP perspectives on child psychiatric diagnostic assessment and treatment planning. J Am Acad Child Adolesc Psychiatry. 2021;60(5):540-542. doi:10.1016/j.jaac.2020.08.470

32. Lee PH, Tervo-Clemmens B, Liu RT, et al. Use of tobacco products and suicide attempts among elementary school-aged children. JAMA Netw Open. 2024;7(2):e240376. doi:10.1001/jamanetworkopen.2024.0376

33. Visoki E, Moore TM, Zhang X, et al. Classification of suicide attempt risk using environmental and lifestyle factors in 3 large youth cohorts. JAMA Psychiatry. 2024;81(10):1020-1029. doi:10.1001/jamapsychiatry.2024.1887

34. Havdahl KA, von Tetzchner S, Huerta M, Lord C, Bishop SL. Utility of the Child Behavior Checklist as a screener for autism spectrum disorder. Autism Res. 2016;9(1):33-42. doi:10.1002/aur.1515

35. Angelakis I, Austin JL, Gooding P. Association of childhood maltreatment with suicide behaviors among young people: a systematic review and meta-analysis. JAMA Netw Open. 2020;3(8):e2012563. doi:10.1001/jamanetworkopen.2020.12563

36. Janiri D, Doucet GE, Pompili M, et al. Risk and protective factors for childhood suicidality: a US population-based study. Lancet Psychiatry. 2020;7(4):317-326. doi:10.1016/S2215-0366(20)30049-3

37. Proust-Lima C, Philipps V, Liquet B. Estimation of extended mixed models using latent classes and latent processes: the R package lcmm. J Stat Softw. 2017;78(2):1-56. doi:10.18637/jss.v078.i02

38. Talbot D, Mésidor M, Chiu Y, Simard M, Sirois C. An alternative perspective on the robust Poisson method for estimating risk or prevalence ratios. Epidemiology. 2023;34(1):1-7. doi:10.1097/EDE.0000000000001544

39. Holmberg MJ, Andersen LW. Estimating risk ratios and risk differences: alternatives to odds ratios. JAMA. 2020;324(11):1098-1099. doi:10.1001/jama.2020.12698

40. Li F, Tong G. Sample size estimation for modified Poisson analysis of cluster randomized trials with a binary outcome. Stat Methods Med Res. 2021;30(5):1288-1305. doi:10.1177/0962280221990415

41. Haneuse S, VanderWeele TJ, Arterburn D. Using the E-value to assess the potential effect of unmeasured confounding in observational studies. JAMA. 2019;321(6):602-603. doi:10.1001/jama.2018.21554

42. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol. 1995;57(1):289-300. doi:10.1111/j.2517-6161.1995.tb02031.x

43. Council on Communications and Media. Children, adolescents, and the media. Pediatrics. 2013;132(5):958-961. doi:10.1542/peds.2013-2656

44. Zablotsky B, Arockiaraj B, Haile G, Ng AE. Daily Screen Time Among Teenagers: United States, July 2021-December 2023. Centers for Disease Control and Prevention; 2024. doi:10.15620/cdc/168509

45. Nagata JM, Cortez CA, Cattle CJ, et al. Screen time use among US adolescents during the COVID-19 pandemic: findings from the Adolescent Brain Cognitive Development (ABCD) Study. JAMA Pediatr. 2022;176(1):94-96. doi:10.1001/jamapediatrics.2021.4334

46. Hedderson MM, Bekelman TA, Li M, et al; Environmental Influences on Child Health Outcomes Program. Trends in screen time use among children during the COVID-19 pandemic, July 2019 through August 2021. JAMA Netw Open. 2023;6(2):e2256157. doi:10.1001/jamanetworkopen.2022.56157

47. Schou Andreassen C, Billieux J, Griffiths MD, et al. The relationship between addictive use of social media and video games and symptoms of psychiatric disorders: a large-scale cross-sectional study. Psychol Addict Behav. 2016;30(2):252-262. doi:10.1037/adb0000160

48. Nagata JM, Singh G, Sajjad OM, et al. Social epidemiology of early adolescent problematic screen use in the United States. Pediatr Res. 2022;92(5):1443-1449. doi:10.1038/s41390-022-02176-8

49. Meshi D, Ellithorpe ME. Problematic social media use and social support received in real-life versus on social media: associations with depression, anxiety and social isolation. Addict Behav. 2021;119:106949. doi:10.1016/j.addbeh.2021.106949

50. Nagata JM, Smith N, Zamora G, et al. Problematic social media use and alcohol expectancies in early adolescents. *BMC Public Health*. 2023;23(1):430. doi:10.1186/s12889-023-15298-1

51. Heeringa SG, Berglund PA. A guide for population-based analysis of the Adolescent Brain Cognitive Development (ABCD) Study baseline data. bioRxiv. Preprint posted online February 10, 2020. doi:10.1101/2020.02.10.942011

52. Viner RM, Gireesh A, Stiglic N, et al. Roles of cyberbullying, sleep, and physical activity in mediating the effects of social media use on mental health and wellbeing among young people in England: a secondary analysis of longitudinal data. *Lancet Child Adolesc Health*. 2019;3(10):685-696. doi:10.1016/S2352-4642(19)30186-5

53. Raney JH, Al-Shoaibi AA, Ganson KT, et al. Associations between adverse childhood experiences and early adolescent problematic screen use in the United States. *BMC Public Health*. 2023;23(1):1213. doi:10.1186/s12889-023-16111-x

54. Raney JH, Testa A, Jackson DB, Ganson KT, Nagata JM. Associations between adverse childhood experiences, adolescent screen time and physical activity during the COVID-19 pandemic. *Acad Pediatr*. 2022;22(8):1294-1299. doi:10.1016/j.acap.2022.07.007

55. Tombeau Cost K, Korczak D, Charach A, et al. Association of parental and contextual stressors with child screen exposure and child screen exposure combined with feeding. *JAMA Netw Open*. 2020;3(2):e1920557. doi:10.1001/jamanetworkopen.2019.20557

56. Ashton JJ, Beattie RM. Screen time in children and adolescents: is there evidence to guide parents and policy? *Lancet Child Adolesc Health*. 2019;3(5):292-294. doi:10.1016/S2352-4642(19)30062-8

57. Nagata JM, Cheng CM, Shim J, et al. Bedtime screen use behaviors and sleep outcomes in early adolescents: a prospective cohort study. *J Adolesc Health*. 2024;75(4):650-655. doi:10.1016/j.jadohealth.2024.06.008

58. Shannon H, Bush K, Villeneuve PJ, Hellemans KG, Guimond S. Problematic social media use in adolescents and young adults: systematic review and meta-analysis. *JMIR Ment Health*. 2022;9(4):e33450. doi:10.2196/33450

59. Xiao Y, Mann JJ, Chow JC, et al. Patterns of social determinants of health and child mental health, cognition, and physical health. *JAMA Pediatr*. 2023;177(12):1294-1305. doi:10.1001/jamapediatrics.2023.4218

60. Choukas-Bradley S, Roberts SR, Maheux AJ, Nesi J. The perfect storm: a developmental-sociocultural framework for the role of social media in adolescent girls' body image concerns and mental health. *Clin Child Fam Psychol Rev*. 2022;25(4):681-701. doi:10.1007/s10567-022-00404-5

Eating and Weight Disorders - Studies on Anorexia, Bulimia and Obesity (2024) 29:57
https://doi.org/10.1007/s40519-024-01685-1

RESEARCH

Screen time, problematic screen use, and eating disorder symptoms among early adolescents: findings from the Adolescent Brain Cognitive Development (ABCD) Study

Jonathan Chu[1] · Kyle T. Ganson[2] · Alexander Testa[3] · Abubakr A. A. Al-shoaibi[1] · Dylan B. Jackson[4] · Rachel F. Rodgers[5,6] · Jinbo He[7] · Fiona C. Baker[8,9] · Jason M. Nagata[1]

Received: 6 December 2023 / Accepted: 21 August 2024

Abstract

Purpose Emerging research evidence suggests positive relationships between higher screen time and eating disorders. However, few studies have examined the prospective associations between screen use and eating disorder symptoms in early adolescents and how problematic screen use may contribute to symptom development.

Methods We analyzed prospective cohort data from the Adolescent Brain Cognitive Development (ABCD) Study ($N = 10{,}246$, 2016–2020, ages 9–14). Logistic regression analyses were used to estimate the longitudinal associations between baseline self-reported screen time and eating disorder symptoms in year two. Logistic regression analyses were also used to estimate cross-sectional associations between problematic screen use in year two (either problematic social media or mobile phone use) and eating disorder symptoms in year two. Eating disorder symptoms based on the Kiddie Schedule for Affective Disorders and Schizophrenia (KSADS-5) included fear of weight gain, self-worth tied to weight, engaging in compensatory behaviors, binge eating, and distress with binge eating.

Results Each additional hour of total screen time and social media use was associated with higher odds of fear of weight gain, self-worth tied to weight, compensatory behaviors to prevent weight gain, binge eating, and distress with binge eating two years later (odds ratio [OR] 1.05–1.55). Both problematic social media and mobile phone use were associated with higher odds of all eating disorder symptoms (OR 1.26–1.82).

Conclusions Findings suggest greater total screen time, social media use, and problematic screen use are associated with more eating disorder symptoms in early adolescence. Clinicians should consider assessing for problem screen use and, when high, screen for disordered eating.

Level of evidence Level III: Evidence obtained from well-designed cohort or case–control analytic studies.

Keywords Eating disorders · Adolescent health · Screen time · Problematic screen use

✉ Jason M. Nagata
jason.nagata@ucsf.edu

1 Department of Pediatrics, University of California, San Francisco, 550 16th Street, 4th Floor, Box 0503, San Francisco, CA 94143, USA

2 Factor-Inwentash Faculty of Social Work, University of Toronto, Toronto, ON, Canada

3 Department of Management, Policy and Community Health, University of Texas Health Science Center at Houston, Houston, TX, USA

4 Department of Population, Family, and Reproductive Health, Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA

5 Department of Applied Psychology, Northeastern University, Boston, MA, USA

6 Department of Psychiatric Emergency and Acute Care, Lapeyronie Hospital, Montpellier, France

7 School of Humanities and Social Science, The Chinese University of Hong Kong, Shenzhen, Guangdong, China

8 Center for Health Sciences, SRI International, Menlo Park, CA, USA

9 School of Physiology, University of the Witwatersrand, Braamfontein, Johannesburg, South Africa

Introduction

Eating disorders are distressing and chronic disorders, linked to significant medical complications and reduced quality of life [1]. Examples of eating disorders include but are not limited to anorexia nervosa, bulimia nervosa, and binge-eating disorder, which are three of the most notable eating disorders among young people around the world [2]. The etiology of eating disorders is thought to be multifactorial. Studies have identified risk factors across biological, psychological, and sociocultural domains, such as genetic predisposition, elevated body mass index (BMI), comorbidity with other mental health disorders, socioeconomic status, and gender [3, 4]. Among these risk factors, a rising area of research is the relationship between screen use, the time spent using devices such as television, video game consoles, and mobile phones for various activities, and eating disorder risk [5, 6].

In recent years, the increasing popularity of social media has led to numerous studies describing the associations between social media use, body image, and eating concerns [7, 8]. However, the majority of these studies have occurred in mostly older female adolescents and young adults (approximately 15–29 years of age). While the focus on this age range is most likely related to the age of onset of clinical diagnoses, studies have shown that eating disorder symptoms may develop in early adolescence [9]. In addition, studies tend to focus on restrictive behaviors, excluding other symptoms of eating disorders such as compensatory behaviors (e.g., vomiting, excessive exercise) and binge eating and also eating disorder cognitions (e.g., feeling self-worth tied to weight, fear of weight gain, and distress with binge eating) that also make up DSM-5 criteria for eating disorder diagnoses [8, 10, 11]. The early adolescent population is known to also have increasing rates of screen use [12], and early symptom development may predispose individuals to long-term disordered eating [3, 13]. Furthermore, early adolescence is a key developmental period in which both the onset of puberty and increased social expectations impact mental health [14]. Therefore, it is imperative to further investigate these risk factors for eating disorders in younger populations to inform advancements in early identification and prevention.

Though the exact mechanisms through which screen time may influence the development of eating disorders is not yet fully understood, the Dual Pathways model describes how pressures to obtain socially constructed body ideals and subsequent body dissatisfaction increase the risk for negative eating disorder cognitions and disordered eating [15]. For example, increased exposure to idealized images of bodies on social media platforms (e.g., Facebook, Instagram, TikTok) may contribute to eating disorder symptoms in youth [16]. One cross-sectional study of 996 Australian adolescents with a mean age of 13 years linked increased social media usage with more disordered eating behaviors, suggesting that these influences may begin at younger ages [17]. However, the cross-sectional design of this study limited its generalizability regarding the directionality of the relationship and thus evidence for social media use as a risk factor for eating disorders. Therefore, longitudinal studies are needed to better understand media use as a potential risk factor for eating disorder symptoms in younger adolescents.

In addition to social media, it is also important to explore how other modalities of screen use, such as television, videos, video games, and texting factor into the potential development of eating disorder symptoms. One prior investigation found longitudinal associations between these screen time modalities and binge-eating disorder in early adolescents [6]. However, the eating outcomes assessed only included binge-eating disorder, and thus, individual symptoms characteristic of anorexia nervosa and bulimia nervosa, such as fearing weight gain, feeling self-worth tied to weight, and inappropriate compensatory behaviors/purging were overlooked. Increased screen time may influence emotional regulation in children and adolescents [5], and prior models have suggested strong associations between emotional dysregulation and eating disorders [18]. Thus, it is of importance to elucidate potential detrimental effects related to eating disorder risk.

Beyond the time spent using screens, further specific screen time behaviors and experiences should be investigated in relation to the risk of developing eating disorder symptoms. For example, problematic social media use, defined as the preoccupation with and compulsion to excessively engage in social media platforms [19], has been linked to deleterious outcomes in physical and mental health, including poorer mental health, sleep disturbances, and dietary problems [5]. Problematic mobile phone use shares similarities with problematic social media use and includes broader applications such as texting, apps, and video chatting. Studies have begun to examine the relationship between problematic screen use and negative eating habits and increased sedentary time [7, 20], suggesting that more problematic screen use is associated with higher body mass index [21, 22].

However, the associations between problematic screen use and eating disorder symptoms (e.g., body dissatisfaction) are less understood. In a large study of adolescents in Slovakia, eating disorder symptoms were associated with excessive internet use and potentially linked to poorer self-control and increased impulsivity [23]. As such, there may exist an overlap between the maladaptive behaviors and symptoms associated with eating disorders and the impulsivity related

to problematic screen use. Additional cross-sectional studies have also shown similar relationships between problematic screen use and eating disorders symptoms, but have primarily been limited to smaller samples and older populations [10, 24, 25]. As the literature has shown that both problematic screen use and eating disorder symptoms may begin in early adolescence [9, 26], further studies are needed to potentially inform early prevention strategies.

The current study aimed to determine the prospective associations between total screen time and social media use at baseline and eating disorder symptoms (e.g., fear of obesity, feeling self-worth is tied to weight, engaging in compensatory behaviors, binge eating, distress with binge eating) at two-year follow-up in a large, national sample of early adolescents. Given the availability of problematic screen use data at the 2-year follow-up, the study also sought to determine the cross-sectional associations between problematic screen use (e.g., problematic social media or mobile phone use) and eating disorder symptoms. To better understand the specific relationship between these screen measures and eating disorder symptoms, we adjusted for potential confounders based on known risk factors, including sociodemographic factors (age, race/ethnicity, household income, parent education status), BMI, anxiety, and impulsivity [3, 26, 27]. We hypothesized that higher total screen time and social media use would be prospectively associated with reporting eating disorder symptoms [6, 16, 17]. We also hypothesized that problematic screen use would be cross-sectionally associated with eating disorder symptoms [24, 25].

Methods

Study population

We analyzed prospective data from the Adolescent Brain Cognitive Development (ABCD) Study, a longitudinal study of brain development and health across adolescence in 11,875 children recruited from 21 sites around the U.S. The ABCD study implemented epidemiologically informed strategies to recruit a sample representative of U.S. diversity, largely through school systems and considering sociodemographic factors. Additional details are described elsewhere [28]. Data analyzed are from the ABCD 4.0 release for the baseline (2016–2018, 9–10 years old), year one (2017–2019) and year two (2018–2020) assessments. Participants with missing data for screen time and eating disorder symptoms were excluded ($N = 1{,}552$, 13.1%, characteristics of included and excluded participants may be found in Additional file 1: Table S1). For participants missing sociodemographic data at baseline, including race/ethnicity, sex, household income, parental education, and study site, we implemented Gaussian normal regression imputation in Stata to impute missing data. Centralized institutional review board (IRB) approval was obtained from the University of California, San Diego. Study sites obtained approval from their respective IRBs. Caregivers provided written informed consent and each child provided written assent. Data used in this study were obtained from the ABCD Study (https://abcdstudy.org), held in the NIMH Data Archive (NDA).

Exposures

Baseline total screen time and social media use

Total screen time and social media use were determined using the self-reported ABCD Youth Screen Time Survey. Participants answered questions about typical hours per day spent on six different screen time modalities (viewing/streaming television shows or movies, watching/streaming videos [e.g., YouTube], playing videogames, texting, video chatting [Skype, Facetime], and social media [e.g., Facebook, Instagram, Twitter]) separately for weekdays and weekend days, based on a previously validated measure [29, 30]. We calculated a weighted average of the participants' typical weekday and weekend screen time use, ((weekday average × 5) + (weekend average × 2))/7, to report a single typical hours per day measure for each modality [22, 31]. We reported the weighted average as a continuous variable after obtaining this screen time total for each modality utilized by participants. Total screen time was determined by summing the weight averages of all modalities.

Year-two problematic screen use

Problematic social media use

Starting in year two, the ABCD Study utilized the adolescent self-reported Social Media Addiction Questionnaire (SMAQ) to assess problematic social media. The six questions of the SMAQ were modeled after the Bergen Facebook Addiction Scale [32]. Examples of the questions included "I've tried to use my social media apps less but I can't" and "I've become stressed or upset if I am not allowed to use my social media apps." Likert-type scale responses ranged from 1 (never) to 6 (very often). Only participants who reported having at least one social media account were asked these items ($n = 5{,}587$).

Problematic mobile phone use

Starting in year two, a similar eight-question Mobile Phone Involvement Questionnaire (MPIQ) was used to assess problematic mobile phone use as reported by adolescents [33]. Examples of questions from the MPIQ included "I interrupt

whatever else I am doing when I am contacted on my phone" and "I lose track of how much I am using my phone." Likert-type scale responses ranged from 1 (strongly disagree) to 7 (strongly agree). This questionnaire has been previously used to assess smartphone dependence and digital multitasking during homework among US high school students [34]. Only participants who reported having mobile phones were asked these items ($n = 7{,}280$).

Outcome: year-two eating disorder symptoms

The ABCD Study utilized the Kiddie Schedule for Affective Disorders and Schizophrenia (KSADS-5), a widely used computerized tool for categorizing child and adolescent mental health concerns based on the DSM-5, for assessment of eating disorder symptoms at two-year follow-up [35, 36]. Participants completed all modules of the KSADS-5 to assess the frequency, duration, and characteristics of eating disorder symptoms. Examples of questions participants were asked included "Do you feel like your self-worth is tied to your weight?" and "Was there ever a time, for a month or longer, that you worried all the time about your weight or becoming fat?" Participants were also asked about behaviors such as compensatory behaviors to lose weight and binge eating. Compensatory behaviors included only eating foods with minimal calories, exercising a lot, throwing up, and taking water pills, laxatives, or diet pills. Those who responded yes to any of the behaviors were coded as engaging in compensatory behaviors to lose weight. Participants were asked about binge eating and whether they experienced distress with binge eating. Additional information regarding the KSADS-5 assessment of eating disorder symptoms used in this study may be found in Additional file 1: Table S2.

Confounders

We selected potential sociodemographic confounders based on previous literature and theory [3, 26, 27, 37]. Age (years), sex (female, male), race/ethnicity (White, Latino/Hispanic, Black, Asian, Native American, other), household income (grouped into two categories reflecting the US median household income: less than $75,000 and $75,000 or more), and highest parent education (high school or less vs. college or more) were based on parents' self-report at baseline. Participant BMI was recorded at baseline. Measures of impulsivity were obtained using the Behavioral Inhibition and Approach Systems scale in the ABCD Study, which assesses participant reward responsiveness, drive, and fun-seeking behavior [38, 39]. Anxiety symptoms at baseline were obtained from parent/caregiver responses to the Child Behavior Checklist (CBCL), a screening tool used to assess psychiatric symptoms and behavior problems in children aged 4–18 [28, 40]. Because participants were asked about eating disorder symptoms at the year two assessment but not asked at baseline, we included parent-reported baseline eating disorder symptoms of their child based on the caregiver KSADS-5 assessment as a confounder in longitudinal analyses. ABCD Study site was included as a confounder to adjust for potential regional variation.

Statistical analysis

Multiple logistic regression analyses were conducted using Stata 18.0 (StataCorp, College Station, TX) to (1) estimate prospective associations between screen time (exposure variable) and the presence of adolescent-reported eating disorder symptoms (fearing obesity, feeling self-worth tied to weight, engaging in compensatory behaviors to lose weight, and binge eating) at two-year follow-up, adjusting for confounders including parent-reported baseline eating disorder symptoms, and (2) estimate cross-sectional associations between problematic screen use and eating disorder symptoms, adjusting for confounders. Additionally, testing for interactions between eating disorder symptoms and sex was not statistically significant, and thus, we did not stratify by sex. Propensity weights developed by the ABCD Study were applied to yield estimates representative of the age, sex, and race/ethnicity distribution of US adolescents based on the American Community Survey from the US Census using the svyset and svy commands in Stata as described in the ABCD Study's guide for population-based analysis [38].

Results

Table 1 describes the sociodemographic characteristics of the 10,246 participants included. The sample was approximately matched by sex (48.6% female) and racially and ethnically diverse (45.6% non-White). At baseline, youth reported an average of 3.9 h of total screen time. At two-year follow-up, 1.4% reported fear of obesity, 1.6% felt their self-worth was tied to their weight, 0.7% engaged in compensatory behaviors to lose weight, 7.5% engaged in binge eating, and 2.9% had distress with binge eating.

Logistic regression analyses examining the prospective associations between baseline screen time and adolescent-reported eating disorder symptoms at two-year follow-up are shown in Table 2. In fully adjusted models, each additional hour of total screen time and social media use was prospectively associated with higher odds of fearing weight gain, feeling self-worth tied to weight, engaging in compensatory behaviors to prevent weight gain, binge eating, and distress with binge eating at two-year follow-up, with odds ratios ranging from 1.05 to 1.55.

Table 3 shows logistic regression analyses examining the cross-sectional associations between problematic screen use

Table 1 Sociodemographic, screen time, problematic screen use, and eating disorder symptoms among 10,246 Adolescent Brain Cognitive Development (ABCD) Study participants

Sociodemographic characteristics (baseline)	Mean (SD)/%
Age (years)	9.9 (0.6)
Sex, *n* (%)	
Female	48.6%
Male	51.4%
Race/ethnicity (%)	
White	54.4%
Latino/Hispanic	19.7%
Black	16.0%
Asian	5.4%
Native American	3.2%
Other	1.4%
Household income (%)	
Less than $75,000	45.0%
$75,000 and greater	55.0%
Parent with college education or more (%)	81.2%
Screen time measures	
Total screen time at baseline (hours)	3.9 (3.1)
Total screen time at year one of follow-up (hours)	4.7 (3.6)
Total screen time at year two of follow-up (hours)	6.1 (5.9)
Social media (hours)	0.1 (0.4)
Problematic screen use measures	
Social media addiction questionnaire score[a]	1.9 (0.9)
Mobile phone involvement questionnaire score[b]	3.1 (1.1)
Eating disorder symptoms	
Fear of obesity	1.4%
Self-worth tied to weight	1.6%
Inappropriate compensatory behaviors to prevent weight gain	0.7%
Binge eating	7.5%
Distress with binge eating	2.9%
BMI (kg/m^2)	18.9 (4.2)
BMI percentile	61.6 (30.8)
Weight (kg)	28.0 (13.5)
Weight percentile	61.8 (29.7)
Anxiety symptoms (t-score)	53.7 (6.3)
BAS reward responsiveness sum score	2.2 (0.6)

Propensity weights were applied to yield representative estimates based on the American Community Survey from the US Census. SD = standard deviation

[a]Asked among a subset who reported social media use ($n = 5{,}587$)

[b]Asked among a subset who reported mobile phone use ($n = 7{,}280$)

(social media use or mobile phone use) and eating disorder symptoms at two-year follow-up. Both problematic social media use and problematic mobile phone use were associated with all eating disorder symptoms in fully adjusted models with odds ratios ranging from 1.26 to 1.82.

Discussion

In this population-based, demographically diverse cohort of early adolescents in the US, we found that greater

Table 2 Associations between baseline total screen time and eating disorder symptoms at two-year follow-up in the Adolescent Brain Cognitive Development Study

Eating disorder symptom	Total screen time[a]		Social media use[a]	
	OR (95% CI)	*p*	OR (95% CI)	*p*
Fear of obesity	**1.12 (1.08–1.17)**	**<0.001**	**1.55 (1.21–1.98)**	**0.001**
Self-worth tied to weight	**1.10 (1.06–1.15)**	**<0.001**	**1.30 (1.03–1.63)**	**0.025**
Inappropriate compensatory behaviors to prevent weight gain	**1.06 (1.03–1.09)**	**<0.001**	**1.18 (1.01–1.40)**	**0.039**
Binge eating	**1.08 (1.05–1.11)**	**<0.001**	**1.28 (1.10–1.49)**	**0.002**
Distress with binge eating	**1.05 (1.01–1.09)**	**0.011**	**1.31 (1.06–1.61)**	**0.012**

Bold indicates $p < 0.05$

[a]Covariates: race/ethnicity, sex, household income, parent education, site, baseline parent-reported eating disorder symptom, baseline BMI percentile, baseline anxiety symptoms, and baseline BAS reward responsiveness

Table 3 Cross-sectional associations between problem screen time use and eating disorder symptoms in the Adolescent Brain Cognitive Development Study

Eating disorder symptom	Problematic social media use[a]		Problematic mobile phone use[a]	
	$n=5{,}587$[b]		$n=7{,}280$[b]	
	OR (95% CI)	*p*	OR (95% CI)	*p*
Fear of obesity	**1.38 (1.11–1.71)**	**0.004**	**1.43 (1.18–1.72)**	**<0.001**
Self-worth tied to weight	**1.75 (1.45–2.10)**	**<0.001**	**1.51 (1.27–1.79)**	**<0.001**
Inappropriate compensatory behaviors to prevent weight gain	**1.43 (1.28–1.60)**	**<0.001**	**1.26 (1.15–1.39)**	**<0.001**
Binge eating	**1.63 (1.48–1.81)**	**<0.001**	**1.66 (1.51–1.81)**	**<0.001**
Distress with binge eating	**1.79 (1.54–2.08)**	**<0.001**	**1.82 (1.57–2.12)**	**<0.001**

Bold indicates $p < 0.05$

[a]Covariates: race/ethnicity, sex, household income, parent education, site, baseline BMI percentile, baseline anxiety symptoms, and baseline BAS reward responsiveness

[b]Assessments for problematic social media and mobile phone use were only performed on participants who responded "yes" to having a social media account or mobile phone, respectively

screen time and social media use were prospectively associated with eating disorder symptoms at two-year follow-up. We also revealed cross-sectional associations between problematic screen use and eating disorder symptoms. In particular, problematic social media use was most strongly associated with feeling self-worth tied to weight, and problematic mobile phone use was most associated with binge eating.

Our findings regarding the relationship between screen time, social media use, and eating disorder symptoms are consistent with prior studies [8, 10, 17]. While this relationship has been previously examined, longitudinal studies are scarce, particularly in younger adolescents, making this an important extension of previous work. Furthermore, as screen time and media use patterns rapidly evolve over time, continued studies are necessary to best capture their potential influence on youth growing up in different periods. Thus, we add to the literature by (1) using a large, national prospective cohort design; (2) focusing on early adolescence, an important period for the development of screen use and eating disorder symptoms; and (3) examining the associations between problematic screen use and eating disorder symptoms.

Of note, social media use only made up a small portion of total screen time in this population of early adolescents and had significant associations particularly with fear of weight gain. Through social media, youth may gain exposure to unrealistic beauty standards that could precipitate low self-esteem, leading to concerns regarding weight and body image [10, 17, 41]. The other forms of screen time that were not focused on in this study and which youth appear to be engaging with at higher amounts (e.g., television, videos, video games, texting) may also expose youth to similar content. Television shows and advertisements frequently depict and glamorize thinness in women and muscularity in men [42]. Influencers across various platforms, such as Instagram, YouTube, or TikTok have been shown to motivate and positively impact people's exercise goals [43]; however, they often portray a "fit" ideal that may similarly lead to body dissatisfaction [20]. Future studies may seek to identify the relationships between

specific screen time modalities and content that place youth at the greatest risk for developing eating disorder symptoms.

The relationship between problematic screen use and disordered eating is less well described in the literature, with existing studies primarily focusing on older adolescents, college students, and young adults [23, 25, 44]. In contrast to benign use, problematic screen use involves dependence and inability to remove oneself from screens, resulting in functional impairment in daily life. Prior studies have shown that problematic screen use and internet addiction may contribute to the development of poor eating habits [45]. For example, individuals may become so engrossed in their screen use that they unwittingly engage in disordered eating behaviors such as skipping meals to spend more time on their devices or binging due to a lack of awareness around how much they have eaten. Some preliminary studies have shown that mindful and intuitive eating practices, approaches to healthy eating that focus on non-judgmental observations of sensations and cognitions during meals, may reduce disordered eating behaviors [46]. As such, it may be possible that the decreased engagement during meals because of problematic screen use can predispose individuals to develop eating habits that then transform into disordered eating.

In our study, the association with the largest odds ratio was between problematic social media use and feeling self-worth tied to weight. Models describing the etiology of eating disorders often include environmental factors such as social pressure regarding physical appearance [47]. Social media has erupted in the last decade, resulting in increased connectedness to peers [48]; however, increased exposure may result in negative cognitions around body dissatisfaction, fearing obesity, and greater emphasis on body image due to social comparisons with content that embodies thinness ideals [10, 17]. Those who engage in problematic social media use are potentially more prone to constantly comparing themselves to other social media users at greater frequencies, which has been shown to have associations with body dissatisfaction and drive for thinness [16]. Consequently, it is possible that constant social media use can make adolescents more vulnerable to these body ideals and feelings of self-worth tied to their weight and body image.

In addition to these negative body image cognitions, we also found that problematic screen use was associated with binge eating and compensatory behaviors to prevent weight gain. Binge eating involves the overconsumption of food in a short period coupled with a loss of control during episodes. In a prior study, we showed that total screen time was longitudinally associated with binge-eating disorder. However, that study did not examine problematic screen use. Combined with purging, which are compensatory behaviors such as vomiting or excessive exercise to prevent weight gain, binge eating also contributes to bulimia nervosa as well as the binge–purge subtype of anorexia nervosa [49]. Theoretical frameworks attempting to explain the etiologies of these disorders have discussed the potential role of impulsivity [50, 51]. The seemingly impulsive nature of binge eating and purging may share similarities with characteristics of addiction and problematic screen use. Poor inhibitory control in impulsivity has well-established links to addictive behaviors [52]. Impulsivity generally refers to taking action or engaging in behaviors without consideration of consequences. High levels of impulsivity are thought to increase the risk of binge–purge episodes and have been demonstrated in longitudinal studies of adults as well as cross-sectional studies of adolescents [51, 53, 54]. Problematic usage and overconsumption of either social media or mobile phones may reflect the similar loss of control and overconsumption exhibited through binge-eating behaviors, which is consistent with our longitudinal findings between total screen time and eating disorder symptoms. Furthermore, children may be prone to overeating in the absence of hunger while distracted in front of screens. Finally, researchers posit that media and advertising content that youth may become exposed to can reflect unattainable body ideals and exacerbate binge eating [41], and adolescents who hold negative feelings towards their own body image are more likely to binge eat [55].

Our study includes notable limitations. Although we adjusted for several potential confounders, including parent-reported baseline eating disorder symptoms, the possibility of residual confounding due to other factors exists. Though the prospective study design for analyses between screen time and eating disorder symptoms improves on prior cross-sectional evidence, we cannot establish causality given the observational nature of the study. As the prevalence of eating disorder symptoms and the diagnoses of eating disorders increase as youth enter later adolescence, additional studies following the ABCD cohort will be an important area of future research. Furthermore, in this study eating disorder symptoms were assessed by parents at baseline and then adolescents at two-year follow-up since participants themselves were not screened for symptoms at baseline. Prior studies have demonstrated parents may provide lower estimates of eating disorder symptoms [56]; however, we acknowledge that generally, there exists discordance between youth–parent reporting of eating disorder pathology that future research may consider evaluating further [56]. Additionally, in this study eating disorder symptoms were analyzed categorically rather than dimensionally, which may not capture the relationship between screen use and the spectrum of symptom severity. It is important to note that the effect sizes of the associations between screen time and eating disorder symptoms were relatively small. However, they are reported for each additional hour, and thus, greater exposure may result in higher odds of developing symptoms. Despite the large sample

size, participants in the study represent adolescents only within the US, which limits generalizability as both screen time and eating disorder patterns can vary in different regions globally [57, 58]. Because problematic screen use measures were not asked at the initial assessment of the ABCD Study, we were unable to determine the prospective associations between problematic screen use and eating disorder symptoms, though this may be another area of future research. Finally, all measures, including evaluations of screen time and eating disorder symptoms, were based on self-reported responses to survey questions and may be subject to reporting bias.

Given the ubiquitous nature of screen and media use in society and the mounting evidence for risks associated with their use, it is imperative to understand their potential downstream effects on youth. Especially with recent rises in both screen use and eating disorders [12, 59], future research should continue to examine their relationship in adolescent populations. Parent education regarding digital media literacy, which has been shown in some studies to decrease screen time in children, can potentially include guidance on body image concerns. The American Academy of Pediatrics encourages the development of Family Media Use Plans, which can include discussions surrounding problematic screen use and disordered eating concerns with children. Clinicians are encouraged to regularly assess screen time in youth, given the accumulating support for its association with a range of poor mental health outcomes. Moreover, clinicians should consider screening for disordered eating in youth who report high or problematic screen use, given the benefits of early identification for prognosis.

Strength and limits

Strengths of the study include the analysis of a large, diverse prospective cohort of early adolescents in the US. Limitations include the use of self-reported measures which could be subject to reporting bias, the lack of problematic screen use measures at baseline, and the possibility of residual confounding.

What is already known on the subject?

Emerging research evidence suggests positive relationships between higher screen time and eating disorders. However, few studies have examined the prospective associations between screen use and eating disorder symptoms in early adolescents and how problematic screen use may contribute to developing eating disorder symptoms.

What does this study add?

Findings suggest greater total screen time, social media use, and problematic screen use are associated with more eating disorder symptoms in early adolescence. Clinicians should consider assessing for problem screen use and, when high, screen for disordered eating.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40519-024-01685-1.

Acknowledgements The authors thank Anthony Kung, William Choi, Seohyeong Lee, and Ashley Saldana for editorial assistance. The ABCD Study was supported by the National Institutes of Health and additional federal partners under award numbers U01DA041022, U01DA041025, U01DA041028, U01DA041048, U01DA041089, U01DA041093, U01DA041106, U01DA041117, U01DA041120, U01DA041134, U01DA041148, U01DA041156, U01DA041174, U24DA041123, and U24DA041147. A full list of supporters is available at https://abcdstudy.org/federal-partners/. A listing of participating sites and a complete listing of the study investigators can be found at https://abcdstudy.org/principal-investigators.html. ABCD consortium investigators designed and implemented the study and/or provided data but did not necessarily participate in the analysis or writing of this report.

Author contributions JC conducted the analysis, drafted the manuscript, and edited the manuscript. KG, AT, DI, RR, JH and FB provided critical revision of the manuscript. JN conceptualized the study, provided critical revision of the manuscript, and provided supervision. All authors approve the final manuscript.

Funding J.M.N. was funded by the National Institutes of Health (K08HL159350 and R01MH135492), the American Heart Association (CDA34760281), and the Doris Duke Charitable Foundation (2022056).

Availability of data and materials Data used in the preparation of this article were obtained from the ABCD Study (https://abcdstudy.org), held in the NIMH Data Archive (NDA).

Declarations

Ethics approval and consent to participate Centralized institutional review board (IRB) approval was obtained from the University of California, San Diego. Study sites obtained approval from their respective IRBs, caregivers provided written informed consent, and each child provided written assent.

Informed consent Not applicable.

Competing interests The authors declare no competing interests.

References

1. Hambleton A, Pepin G, Le A, Maloney D, National Eating Disorder Research Consortium, Aouad P, et al (2022) Psychiatric and medical comorbidities of eating disorders: findings from a rapid review of the literature. J Eat Disord 10:132
2. Silén Y, Keski-Rahkonen A (2022) Worldwide prevalence of DSM-5 eating disorders among young people. Curr Opin Psychiatry 35:362–371
3. Striegel-Moore RH, Bulik CM (2007) Risk factors for eating disorders. Am Psychol 62:181–198
4. Barakat S, McLean SA, Bryant E, Le A, Marks P, National Eating Disorder Research Consortium, et al (2023) Risk factors for eating disorders: findings from a rapid review. J Eat Disord 11:8
5. Lissak G (2018) Adverse physiological and psychological effects of screen time on children and adolescents: literature review and case study. Environ Res 164:149–157
6. Nagata JM, Iyer P, Chu J, Baker FC, Pettee Gabriel K, Garber AK, et al (2021) Contemporary screen time modalities among children 9–10 years old and binge-eating disorder at one-year follow-up: a prospective cohort study. Int J Eat Disord 54:887
7. Holland G, Tiggemann M (2016) A systematic review of the impact of the use of social networking sites on body image and disordered eating outcomes. Body Image 17:100–110
8. Fardouly J, Vartanian LR (2016) Social media and body image concerns: current research and future directions. Curr Opin Psychol 9:1–5
9. Murray SB, Ganson KT, Chu J, Jann K, Nagata JM (2022) The prevalence of preadolescent eating disorders in the United States. J Adolesc Health 70:825–828
10. Sidani JE, Shensa A, Hoffman B, Hanmer J, Primack BA (2016) The association between social media use and eating concerns among US young adults. J Acad Nutr Diet 116:1465
11. Sarmiento C, Lau C (2020) Diagnostic and statistical manual of mental disorders, 5th ed. In: The Wiley encyclopedia of personality and individual differences. DSM-5
12. Nagata JM, Cortez CA, Cattle CJ, Ganson KT, Iyer P, Bibbins-Domingo K, et al (2022) Screen time use among US adolescents during the COVID-19 pandemic: findings from the adolescent brain cognitive development (ABCD) Study. JAMA Pediatr 176:94
13. Littleton HL, Ollendick T (2003) Negative body image and disordered eating behavior in children and adolescents: what places youth at risk and how can these problems be prevented? Clin Child Fam Psychol Rev 6:51–66
14. Dorn LD, Hostinar CE, Susman EJ, Pervanidou P (2019) Conceptualizing puberty as a window of opportunity for impacting health and well-being across the life span. J Res Adolesc 29:155
15. Maraldo TM, Zhou W, Dowling J, Vander Wal JS (2016) Replication and extension of the dual pathway model of disordered eating: the role of fear of negative evaluation, suggestibility, rumination, and self-compassion. Eat Behav 23:187–194
16. Lonergan AR, Bussey K, Fardouly J, Griffiths S, Murray SB, Hay P, et al (2020) Protect me from my selfie: Examining the association between photo-based social media behaviors and self-reported eating disorders in adolescence. Int J Eat Disord 53:755
17. Wilksch SM, O'Shea A, Ho P, Byrne S, Wade TD (2020) The relationship between social media use and disordered eating in young adolescents. Int J Eat Disord 53:96
18. Paulus FW, Ohmann S, Möhler E, Plener P, Popow C (2021) Emotional dysregulation in children and adolescents with psychiatric disorders. a narrative review. Front Psychiatry 12:628252
19. Domoff SE, Borgen AL, Radesky JS (2020) Interactional theory of childhood problematic media use. Hum Behav Emerg Technol 2:343–353
20. Holland G, Tiggemann M (2017) "Strong beats skinny every time": Disordered eating and compulsive exercise in women who post fitspiration on Instagram: disordered eating in women who post fitspiration. Int J Eat Disord 50:76–79
21. Julian V, Bergsten P, Forslund A, Ahlstrom H, Ciba I, Dahlbom M, et al (2022) Sedentary time has a stronger impact on metabolic health than moderate to vigorous physical activity in adolescents with obesity: a cross-sectional analysis of the Beta-JUDO study. Pediatr Obes. 17:e12897
22. Nagata JM, Iyer P, Chu J, Baker FC, Gabriel KP, Garber AK, et al (2021) Contemporary screen time usage among children 9–10-years-old is associated with higher body mass index percentile at 1-year follow-up: a prospective cohort study. Pediatr Obes 16:e12827
23. Šablatúrová N, Gottfried J, Blinka L, Ševčíková A, Husarova D (2021) Eating disorders symptoms and excessive internet use in adolescents: the role of internalising and externalising problems. J Eat Disord 9:152
24. Panea-Pizarro I, López-Espuela F, Martos-Sánchez A, Domínguez-Martín AT, Beato-Fernández L, Moran-Garcia JM (2020) Internet addiction and Facebook addiction in Spanish women with eating disorders. Arch Psychiatr Nurs 34:442
25. Tayhan Kartal F, Yabancı Ayhan N (2021) Relationship between eating disorders and internet and smartphone addiction in college students. Eat Weight Disord 26:1853
26. Nagata JM, Singh G, Sajjad OM, Ganson KT, Testa A, Jackson DB et al (2022) Social epidemiology of early adolescent problematic screen use in the United States. Pediatr Res 92:1443–1449
27. Nagata JM, Ganson KT, Iyer P, Chu J, Baker FC, Pettee Gabriel K et al (2022) Sociodemographic correlates of contemporary screen time use among 9- and 10-year-old children. J Pediatr 240:213-220.e2
28. Barch DM, Albaugh MD, Avenevoli S, Chang L, Clark DB, Glantz MD, et al (2018) Demographic, physical and mental health assessments in the adolescent brain and cognitive development study: rationale and description. Dev Cogn Neurosci 32:55
29. Bagot KS, Matthews SA, Mason M, Squeglia LM, Fowler J, Gray K et al (2018) Current, future and potential use of mobile and wearable technologies and social media data in the ABCD study to increase understanding of contributors to child health. Dev Cogn Neurosci 32:121–129
30. Sharif I, Wills TA, Sargent JD (2010) Effect of visual media use on school performance: a prospective study. J Adolesc Health 46:52–61
31. Guerrero MD, Barnes JD, Chaput JP, Tremblay MS (2019) Screen time and problem behaviors in children: exploring the mediating role of sleep duration. Int J Behav Nutr Phys Act 16:105
32. Andreassen CS, Torsheim T, Brunborg GS, Pallesen S (2012) Development of a Facebook Addiction Scale. Psychol Rep 110:501–517
33. Walsh SP, White KM, McD YR (2010) Needing to connect: the effect of self and others on young people's involvement with their mobile phones. Aust J Psychol 62:194–203
34. Mrazek AJ, Mrazek MD, Ortega JR, Ji RR, Karimi SS, Brown CS et al (2021) Teenagers' smartphone use during homework: an analysis of beliefs and behaviors around digital multitasking. Educ Sci 11:713
35. Townsend L, Kobak K, Kearney C, Milham M, Andreotti C, Escalera J et al (2020) Development of three web-based computerized versions of the kiddie schedule for affective disorders and

schizophrenia child psychiatric diagnostic interview: preliminary validity data. J Am Acad Child Adolesc Psychiatry 59:309–325
36. Cheng CM, Chu J, Ganson KT, Trompeter N, Testa A, Jackson DB et al (2023) Cyberbullying and eating disorder symptoms in US early adolescents. Int J Eat Disord 56:2336–2342
37. Nagata JM, Smith-Russack Z, Paul A, Saldana GA, Shao IY, Al-Shoaibi AAA et al (2023) The social epidemiology of binge-eating disorder and behaviors in early adolescents. J Eat Disord 11:182
38. Heeringa SG, Berglund PA (2020) A guide for population-based analysis of the adolescent brain cognitive development (ABCD) study baseline data. bioRxiv. https://doi.org/10.1101/2020.02.10.942011
39. Pagliaccio D, Luking KR, Anokhin AP, Gotlib IH, Hayden EP, Olino TM et al (2016) Revising the BIS/BAS Scale to study development: measurement invariance and normative effects of age and sex from childhood through adulthood. Psychol Assess 28:429–442
40. Achenbach TM, Ruffle TM (2000) The child behavior checklist and related forms for assessing behavioral/emotional problems and competencies. Pediatr Rev 21:265–271
41. Aparicio-Martinez P, Perea-Moreno A-J, Martinez-Jimenez MP, Redel-Macías MD, Pagliari C, Vaquero-Abellan M (2019) Social media, thin-ideal, body dissatisfaction and disordered eating attitudes: an exploratory analysis. Int J Environ Res Public Health 16:4177
42. Tiggemann M, Pickering AS (1996) Role of television in adolescent women's body dissatisfaction and drive for thinness. Int J Eat Disord 20:199–203
43. Li W, Ding H, Xu G, Yang J (2023) The impact of fitness influencers on a social media platform on exercise intention during the COVID-19 pandemic: the role of parasocial relationships. Int J Environ Res Public Health 20:1113
44. Hinojo-Lucena FJ, Aznar-Díaz I, Cáceres-Reche MP, Trujillo-Torres JM, Romero-Rodríguez JM (2019) Problematic Internet use as a predictor of eating disorders in students: a systematic review and meta-analysis study. Nutrients 11:2151
45. Kim Y, Park JY, Kim SB, Jung I-K, Lim YS, Kim J-H (2010) The effects of Internet addiction on the lifestyle and dietary behavior of Korean adolescents. Nutr Res Pract 4:51
46. Anderson LM, Reilly EE, Schaumberg K, Dmochowski S, Anderson DA (2016) Contributions of mindful eating, intuitive eating, and restraint to BMI, disordered eating, and meal consumption in college students. Eat Weight Disord Stud Anorex Bulim Obes 21:83–90
47. Rikani AA, Choudhry Z, Maqsood Choudhry A, Ikram H, Waheed Asghar M, Kajal D et al (2013) A critique of the literature on etiology of eating disorders. Ann Neurosci 20:157–161
48. Ryan T, Allen KA, Gray DL, McInerney DM (2017) How social are social media? A review of online social behaviour and connectedness. J Relatsh Res 8:e8
49. American Psychiatric Association (2013) Diagnostic and statistical manual of mental disorders. American Psychiatric Publishing, Arlington, VA
50. Lavender JM, Mitchell JE (2015) Eating disorders and their relationship to impulsivity. Curr Treat Options Psychiatry 2:394–401
51. Wonderlich SA, Connolly KM, Stice E (2004) Impulsivity as a risk factor for eating disorder behavior: assessment implications with adolescents. Int J Eat Disord 36:172–182
52. Lee RSC, Hoppenbrouwers S, Franken I (2019) A systematic meta-review of impulsivity and compulsivity in addictive behaviors. Neuropsychol Rev 29:14–26
53. Claes L, Vandereycken W, Vertommen H (2005) Impulsivity-related traits in eating disorder patients. Personal Individ Differ 39:739–749
54. Fischer S, Peterson CM, McCarthy D (2013) A prospective test of the influence of negative urgency and expectancies on binge eating and purging. Psychol Addict Behav 27:294–300
55. Lewer M, Bauer A, Hartmann A, Vocks S (2017) Different facets of body image disturbance in binge eating disorder: a review. Nutrients 9:1294
56. Mariano P, Watson HJ, Leach DJ, McCormack J, Forbes DA (2013) Parent–child concordance in reporting of child eating disorder pathology as assessed by the eating disorder examination. Int J Eat Disord 46:617–625
57. Hoek HW (2016) Review of the worldwide epidemiology of eating disorders. Curr Opin Psychiatry 29:336–339
58. Mullan K, Hofferth SL (2022) A comparative time-diary analysis of UK and US children's screen time and device use. Child Indic Res 15:795–818
59. Rodgers RF, Lombardo C, Cerolini S, Franko DL, Omori M, Fuller-Tyszkiewicz M, et al (2020) The impact of the COVID-19 pandemic on eating disorder risk and symptoms. Int J Eat Disord 53:1166

International Journal of
*Environmental Research
and Public Health*

MDPI

Systematic Review

Time Spent on Social Media and Risk of Depression in Adolescents: A Dose–Response Meta-Analysis

Mingli Liu [1,*], Kimberly E. Kamper-DeMarco [2], Jie Zhang [3,4], Jia Xiao [1], Daifeng Dong [5,6] and Peng Xue [7]

1 Department of Psychology, Hunan University of Science and Technology, Xiangtan 411201, China; xiaojia2019995@163.com
2 Department of Psychology, State University of New York College at Buffalo, Buffalo, NY 14222, USA; kamperke@buffalostate.edu
3 Department of Sociology, Central University of Finance and Economics, Beijing 102206, China; zhangj@buffalostate.edu
4 Department of Sociology, State University of New York College at Buffalo, Buffalo, NY 14222, USA
5 Medical Psychological Center, The Second Xiangya Hospital of Central South University, Changsha 410011, China; daifengdong@csu.edu.cn
6 China National Clinical Research Center for Mental Disorders (Xiangya), Changsha 410011, China
7 College of Psychology, Liaoning Normal University, Dalian 116082, China; xuepeng486@163.com
* Correspondence: liumingli@hnust.edu.cn

Abstract: Adolescent depression is a worldwide public health concern and has contributed to significant socioeconomic burden. Investigating the association between time spent on social media (TSSM) and depression may provide guidance toward the prevention and intervention of adolescent depression. However, related literature reported mixed findings in terms of the relationship between TSSM and depression in adolescents. Hence, we conducted a comprehensive dose–response meta-analysis to clarify this issue. We conducted a systematic title/abstract and topic search of the relative terms in Web of Science, PubMed, PsycINFO databases through 9 January 2022. Odd ratios (ORs) were used to examine the pooled effect size of the association between TSSM and risk of depression. Dose-response analysis was evaluated by a generalized least squares trend estimation. Twenty-one cross-sectional studies and five longitudinal studies including a total of 55,340 participants were included. Overall, more TSSM was significantly associated with a higher risk of depression symptoms (OR = 1.60, 95%CI: 1.45 to 1.75) with high heterogeneity (Q_{29} = 105.9, $p < 0.001$; I^2 = 72.6%). The association was stronger for adolescent girls (OR = 1.72, 95%CI: 1.41 to 2.09) than boys (OR = 1.20, 95%CI: 1.05 to 1.37). Five studies with seven reports were included in dose-response analysis. There was a linear dose-response association of TSSM and risk of depression. The risk of depression increased by 13% (OR = 1.13, 95%CI: 1.09 to 1.17, $p < 0.001$) for each hour increase in social media use in adolescents. TSSM is associated with depression in a linear dose-response and gender-specific manner, which suggests the need for better monitoring of adolescent social media use. However, motivation, content, and engagement on and exposure to social media use may also be important contributing factors, making it necessary to interpret the current findings with caution. Therefore, further research is required to clarify not only the causal link between TSSM and depression by randomized control studies but also the influence of other factors, such as active vs. passive social media use or different types of engagement or environments in which social media is used.

Keywords: social media use; depression; adolescents; meta-analysis; dose-response

check for updates

Citation: Liu, M.; Kamper-DeMarco, K.E.; Zhang, J.; Xiao, J.; Dong, D.; Xue, P. Time Spent on Social Media and Risk of Depression in Adolescents: A Dose-Response Meta-Analysis. *Int. J. Environ. Res. Public Health* **2022**, *19*, 5164. https://doi.org/10.3390/ijerph19095164

Academic Editors: Peter Choate, Christina Tortorelli, Carly McMorris and Daniel Kikulwe

Received: 17 March 2022
Accepted: 19 April 2022
Published: 24 April 2022

1. Introduction

Social media, also known as social networking, are internet-based interactive platforms where individuals and communities share and communicate [1,2]. In society today, children and adolescents grow up having both in-person and virtual social connections through social media (e.g., Facebook, Instagram, and WeChat) [3]. This continually emerging

internet-based social communication has greatly expanded adolescents' ability to make friends worldwide and makes it possible to connect with others anytime and anywhere. However, there is an ongoing debate about whether social media use is harmful to mental health or not [4,5], with some prior findings highlighting the psychological risk, especially depression, associated with excessive time spent on social medial (TSSM) in adolescence [6], while other studies report that there are only circumstantial correlations between TSSM and psychological problems [7]. Therefore, whether TSSM is associated with adolescents' mental health concerns is still unclear. Notably, an increase in depression has emerged in adolescence, particularly in adolescent girls, over the past ten years [4]. Social media use has also been increasing rapidly at the same time [3,6]. Thus, it is necessary to obtain insight into the association between TSSM and depression in adolescents.

Several theories may explain the inconsistent findings regarding TSSM and depression in adolescents. Based on both the uses and gratifications theory [8] and self-determination theory [9], adolescents may gain a sense of belonging [10,11] and increased self-esteem [12,13] through social media, which is then associated with lower levels of depression. In support of these theories, the association of TSSM and depression in adolescents would follow a U-shaped curve as a previous work suggested [14]. One study determined that the lowest risk of depression in adolescence was found when individuals use approximately 1 h of screen time per day compared with the no screen time group. Some work has also demonstrated the benefits of social media use on depression risk, while other studies have reported very small or null correlations between the two [7,15–18]. Currently, there are more studies reporting a significant positive association between TSSM and depression in adolescents [6,19–22], with a recent study supported a J-shaped curve between TSSM and depression [5]. The displacement hypothesis [23] may help to explain the dark side of excessive TSSM on depression. According to this hypothesis, TSSM may replace time for productive and/or active activities, such as physical activity or face-to-face interpersonal communication, thereby influencing adolescents' overall mental health, including depressive symptoms. Meanwhile, the strain theory may also explain the link between excessive TSSM and higher depression. Strains are usually caused by negative life events [24]. In a heavy involvement in TSSM, adolescents may experience more value strain, aspiration strain, and deprivation strain. All those strains are more likely to lead to depression [25]. Undoubtedly, another way to account for the high correlation between the two variables is to say that depressed adolescents may be more likely to indulge in social media to kill time [26] because depressed individuals have a negative cognitive bias, which may impair adolescents' self-regulation and result in excessive social media use. Finally, these contradictory findings regarding social media use and adolescent depression may also be related to the different methodologies that these studies used, such as the use of different populations and measurements.

Another noteworthy variable is gender. Most studies reported mixed-gender results about social media use and depression. Recently, researchers found that associations between TSSM and depression are different in boys and girls, with TSSM only associated with depression for girls [19,27]. This is in line with research demonstrating that girls use much more social media and place more importance on the closeness of their interpersonal relationships than boys [5], which may then lead to more relational aggression, fear of missing out on social media, and depression [28,29]. It should be noted that one study did find that boys' TSSM was also associated with depression; however, there was a stronger correlation for girls [6]. It is important to note that the relationship between TSSM and depression across gender is far more complicated than is outlined above. Numerous other factors may also affect these associations, such as active or passive use of social media [30], motivations for use [31], or environments to which adolescents are exposed [32].

To clarify the association between social media use and risk of depression in adolescents, several reviews have qualitatively summarized this association in children and adolescents [3,30,33,34]; however, these reviews often lack quantitative assessments. Relatedly, in these reviews, social media use is often measured more broadly, including other

information, such as frequency, purpose, and investment or addiction of social media use, instead of specifically TSSM. One prior meta-analysis [35] pooled the correlation of social media use (not specifically time-based) and depression in adolescents from 11 studies and found a small but statistically significant positive correlation with high heterogeneity. In another meta-analysis, social media use was measured using both TSSM and frequency of social media use, and similar findings were reported [36]. However, responses regarding frequency of social media use were most commonly measured on a scale from "never" to "almost every day", which have little variability because most adolescents use social media every day [5]. Similarly, neither meta-analysis explored sources of heterogeneity, and there was no pooled estimation of the relation between TSSM and risk of depression for adolescents, which may be important for providing evidence-based guidelines regarding TSSM. Thus far, there has been no meta-analysis that has examined a dose–response association between TSSM and risk of depression. In sum, it is clear that a more comprehensive meta-analysis is needed to quantify the dose–response association between TSSM and the risk of depression in adolescents.

The purpose of the current study is to summarize evidence related to the association between TSSM and depression in adolescents by pooling the risk of depression with TSSM for adolescents, quantifying a dose–response association, and exploring the heterogeneity of the included studies. Based on displacement theory, we hypothesize that more TSSM will be associated with a higher risk of depression in adolescents, with a linear dose–response association. Moderation by gender will also be included as an exploratory hypothesis. However, because of its non-experimental nature, no causal inferences can be drawn in the current study.

2. Methods

2.1. Search Strategy

This study was conducted according to the PRISMA guidelines [37] (see Table S1). Electronic databases, including Web of Science, PubMed, and PsycINFO, were searched systematically using title/abstract, and topic (through 9 January 2022) with no publication type or language restriction. To determine "social media" related search terms, a stratified searching strategy was adopted. Firstly, we searched the PubMed database using the most general terms of social media, such as "social media", "digital media", "social networking", "SNS", and "screen media". Next, we screened all study titles and hundreds of abstracts, ending up with 56 different terms. Based on using frequency and generality, we divided them into two categories: general social media-related terms and specific social media-related terms. All general social media terms were included in the final search. However, some rarely used specific social media-related terms, such as "Digg", and "Edmodo", were deleted. Finally, three sets of medical subject terms (MeSH) and their combinations were used in the search, including "social media", "social network*", "SNS", "digital media", "screen media", "online media", "internet media", "collaborative filtering site*", "media sharing site*", "Mashups", "Facebook", "Twitter", "Instagram", "YouTube", "Snapchat", "LinkedIn", "WhatsApp", "Pinterest", "Blog", "Wiki", "Tumblr", "Myspace", "Google+", "Reddit", "WeChat", "QQ", "WordPress", "Telegram", "Flickr", "Skype", "Vine", "Tweeting", "podcasts", "Tik Tok", "Sermo", "Google Groups", "Forum and Blog", "Second Life"; "depress*"; "adolescen*", "juvenile*", "teenager*", "high school student*", "middle school student*", "children". The asterisk indicates that the search was inclusive of larger words that contained the word or word fragment. Additionally, references of retrieved articles were screened.

2.2. Inclusion and Exclusion Criteria

Studies were included if the following criteria were fulfilled: they were an observational study; they reported complete correlation indices of TSSM with depression which could be subsequently converted into an odds ratio (OR) with 95%CI; and average participant age was between 10 and 19 years old. Articles not meeting the inclusion criteria were

excluded. Studies were also excluded if they reported mixed screen time, such as time spent playing internet games or watching online videos, as this would cause the measure of TSSM to be unclear. Authors were contacted if data were missing. Only one study was included if multiple articles reported the same research. The screening of titles/abstracts and topic and subsequent full-text assessment were performed independently by two authors (J.X. and P.X.). When the two authors made different decisions, they discussed the full text and determine its eligibility for inclusion together. If the two authors still disagreed, a third author (M.L.) helped to resolve the disagreement. Figure 1 displays the screening process.

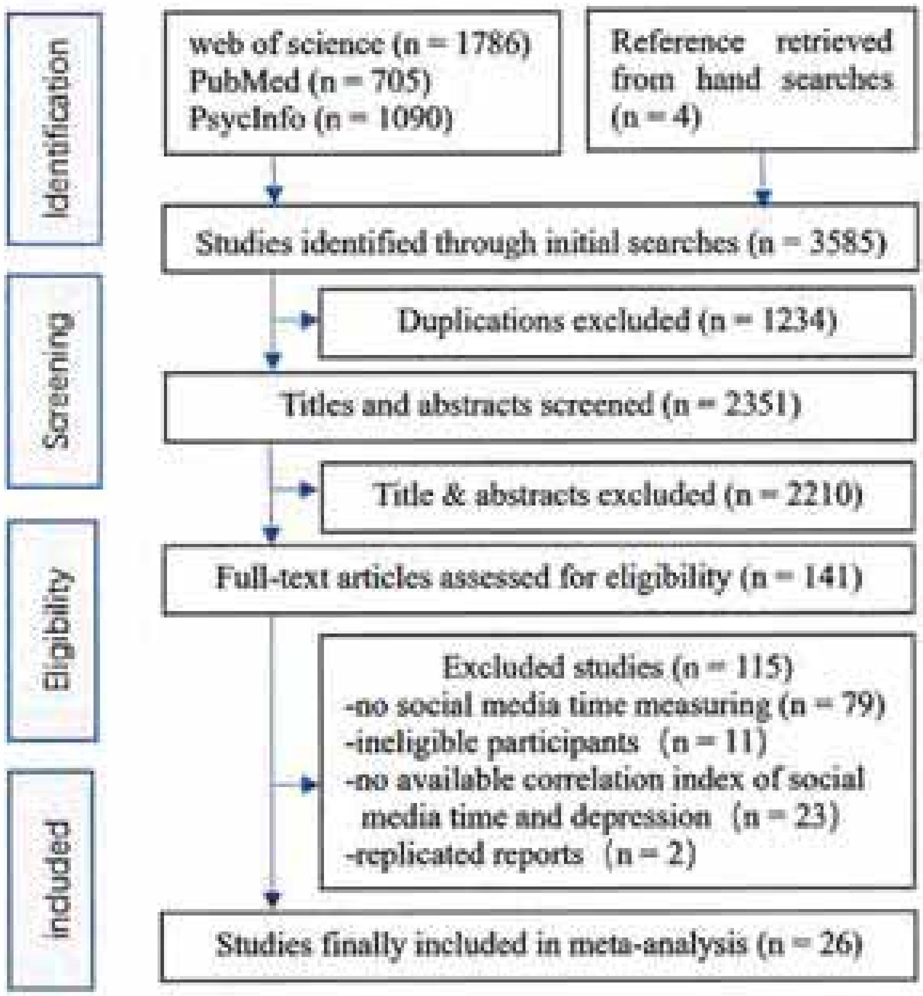


Figure 1. Flow chart of article screening process.

2.3. Data Extraction

All related data (i.e., the first author's name, published year, country, study objective, study design, participants' gender and age, sample size, number of cases (for dose–response analyses), the detailed measure of TSSM and depression, and the correlation index of TSSM with depression) of eligible studies were extracted using EpiData V3.1 and Excel by two investigators. For the quality assessment, we referenced the Meta-analysis of Observational Studies in Epidemiology (MOOSE) [38] and the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) [39] guidelines. Study quality was rated on a scale with a maximum of 8 points based on the following criteria: appropriate selection of participants (1 point); proper measures of TSSM (2 points) and depression (2 points);

appropriate methods to deal with the design issues (1 point); appropriate handling of confounds (1 point) and proper statistical methods (1 point).

2.4. Statistical Analyses

Pooled data were expressed as ORs with 95%CIs. Studies that provided effect sizes stratified by gender were treated as two separate reports. For the studies reporting correlation coefficients, we converted the correlation coefficients to ORs with 95%CIs [40]. For one study using 1–3 h/day as the reference category for TSSM, we recalculated the ORs and 95%CIs using the no TSSM group as the reference category [40]. For studies that provided multiple categories' effect sizes, we combined the corresponding estimates using the Excel RRs proposed by Hamling et al. [41]. The Q statistic was used to evaluate the heterogeneity among studies, and it was quantified by I^2. Low, moderate, and high heterogeneity were indicated by the 25%, 50%, and 75% values of I^2, respectively. If $I^2 < 50\%$, a fixed-effects model was used to estimate the pooled OR and corresponding 95%CI; otherwise, a random-effects model was administrated. To assess the sources of heterogeneity, we performed several subgroup analyses including gender, geographical regions, the measure of TSSM and depression, and sample size. In addition, sensitivity analyses were conducted to test the robustness of the results. Furthermore, funnel plot asymmetry was used to detect publication bias of the included studies in this meta-analysis, and then Begg's and Egger's tests were performed to measure the publication bias.

A specific dose–response analysis was conducted to further estimate the association between TSSM and risk of depression. For the studies that did not report the median or mean of each category, the dose was calculated as the midpoint of the lower and upper boundaries in each group; for the open-ended lower or upper group, the boundary was assumed as the same as the closest group. Both non-linear and linear associations between TSSM and depression were tested. The potential non-linear dose–response relationship between TSSM and depression was estimated using a restricted cubic spline model with three knots of the TSSM distribution. Significance was then tested by setting the second spline coefficient equal to zero. A random-effects model was conducted to examine the trend because of the high heterogeneity among the studies. The dose–response coefficients and corresponding 95%CIs were calculated using a generalized least squares regression. The significance level was set at $p < 0.05$. All statistical analyses of this study were performed with STATA V12 software (Stata Corp, College Station, TX, USA).

3. Results

3.1. Characteristics of the Included Studies

According to the inclusion and exclusion criteria, 30 reports from a total of 26 studies, including a total of 55,340 participants, were included in the final analyses (see Figure 1). The characteristics of the included studies are summarized in Table 1. Twenty-one studies [6,15–22,27,42–51] were cross-sectional, and five were longitudinal [52–56]. Of note, one longitudinal study conducted by Coyne et al. [27] was actually a cross-sectional design for the relationship between TSSM and depression because they reported eight cross-sectional correlations based on data collected from each wave in eight years (2009 to 2017). We incorporated the seventh wave data, which was conducted in the most recent year and which also met the age criteria (19 years). For another four-wave longitudinal study [56], the authors reported a general between-persons and a within-persons regression association with means and standard deviations in the first wave and the last wave. We converted the data into ORs with 95%CIs using the general between-persons correlation for the four waves, with means and standard deviations in the last wave. Across all studies, sample size varied widely from 85 to 11,423 participants. Five studies [6,19,27,49,51] analyzed gender groups separately. However, we combined the total effect size for one study [51] that reported gender-specific results because the sample sizes of the single genders were too small. The mean age of all participants ranged from 11 to 19 years. Three studies reported the age range of the participants with no exact mean ages provided [21,42,47].

Eleven studies were conducted in Europe [6,16,18,19,21,43,47,49,52,53], nine in North America [27,31,42,44,46,48,50,51,56], four in Asia [15,17,45,55], one in Brazil [20], and one in Australia [22]. For the measure of TSSM, most of the studies (22 of 26) used total TSSM while the other four studies [31,43,46,53] used time spent on specific social media platforms, such as Facebook or Instagram. Meanwhile, several questionnaires, including the Center for Epidemiological Studies-Depression scale [57] (CESD, 11 studies), the Short version of the Mood and Feelings Questionnaire (SMFQ, 7 studies) [58], the Beck Depression Inventory [59] (BDI, 2 studies), the Patient Health Questionnaire-9 [60] (PHQ9, 3 studies), the Children's Depression Inventory [61] (CDI, 1 study), the Brief Symptom Inventory [62] (BSI, 3 studies), the Hospital Anxiety and Depression Scale [63] (HADS, 1 study), the scale of the Original Symptom Checklist-Depression dimension [64] (OSCD, 1study), and one question asking "how often you felt depressed" [54] were used to measure depressive symptoms across all included studies. The quality score of all included studies ranged from 3 to 7, with 19 studies obtaining a score of greater than 5 (see Table S2).

3.2. Associations between TSSM and Depression Risk

The overall pooled OR was 1.59 (95%CI: 1.44 to 1.77; $p < 0.001$) with high heterogeneity ($Q_{(27)} = 105.9$, $p < 0.001$; $I^2 = 72.6\%$). The combined OR was 1.61 (95%CI: 1.44 to 1.81) with high heterogeneity ($Q_{(24)} = 97.25$, $I^2 = 75.3\%$) for cross-sectional studies and 1.57 (95%CI: 1.44 to 1.71) with almost zero heterogeneity ($Q_{(4)} = 3.46$, $I^2 = 0\%$) for longitudinal studies (see Figure 2).

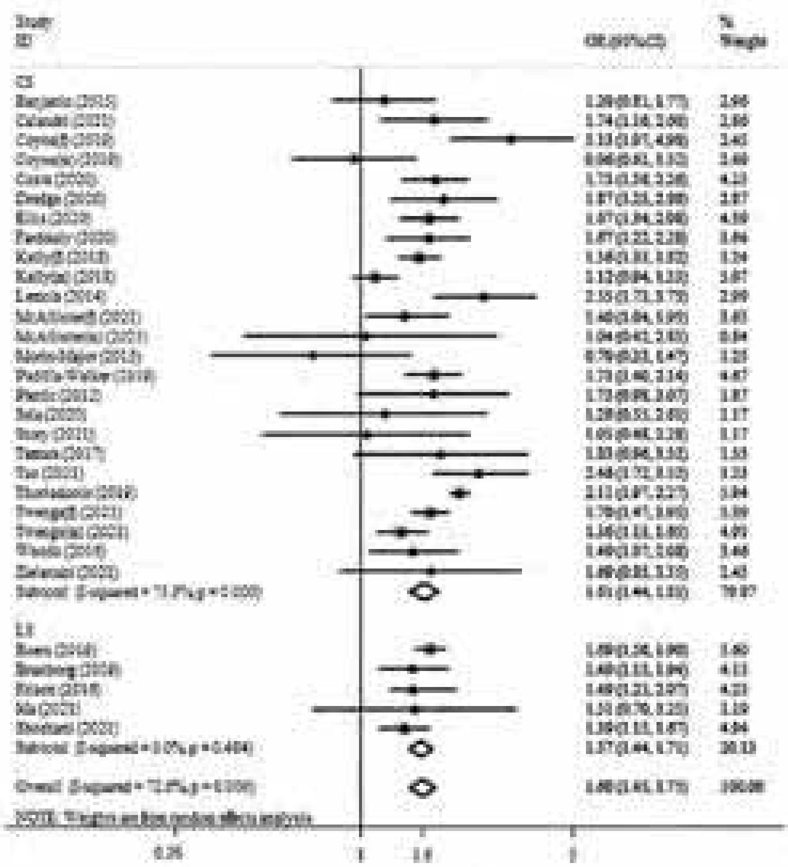


Figure 2. Forest plot of the association between time spent on social media (hours/day) and risk of depression in adolescents by study design. OR of depression for higher daily time using social media compared with reference groups and corresponding 95%CI. CS, cross-sectional; LS, longitudinal; f, female; m, male.

Table 1. The Characteristics of the included studies.

Study	Design	Main Study Objective	Country; Sample Size (Female)	Age (Years)	Measure of Time Spent on Social Media	Depression Measure
Banjanin et al. 2015	CS	Investigated the potential relationship between internet addiction and depression in adolescents.	Serbia; 336 (66%)	18	Self-report daily time spent on social networking; Response: self-administered open answer	CESD
Boers et al. 2019	LS	Repeatedly measured the association between screen time and depression.	Canada; 3826 (47%)	12.7–15.7 Grade 7–11	Self-report how much time per day they spend on social networking sites; Response: 0–30 min, 30 min–1.5 h, 1.5 h–2.5 h, ≥3.5 h	BSI
Brunborg et al. 2019	LS	Examined association between time spent on social media and depression, conduct problems, and drinking.	Norway; 763 (55%)	15.22	Self-report daily hours spent on social media; Response: <1 to >15 in hourly increments	PHQ9
Calandri et al. 2021	LS	Investigated the relationships between social media use and depressive symptoms.	Italy; 336 (48%)	13.0 (13–15)	Self-report daily hrs spent on communicating online with friends through social networks; Response: 0, 1, 2, ≥3	CESD
Costa et al. 2020	CS	Examined the associations between self-reported and accelerometer-measured movement behaviors and depressive symptoms.	Brazil; 610 (52%)	16.30 (14–18)	Self-report daily hours spent on social media; Response: <2, 2–4, ≥4	CESD
Coyne et al. 2019	CS	Examined the association between time spent using social media and depression and anxiety at the intra-individual level.	USA; 500 (52%)	13–20	Self-report daily hours on social media; Response: 1 (0) to 9 (>8)	CESD
Dredge et al. 2020	CS	Examined the association between online gaming and social media use frequency, depression, and other mental health.	China; 320 (47%)	13.98 (12–17)	Self-report daily time spent on social media; Response: 1 (0) to 9 (>8)	PHQ9
Ellis et al. 2020	CS	Examined the relationships between psychological adjustment and stress and the initial COVID-19 crisis.	Canada; 1054 (76%)	16.68 (14–18)	Self-report daily time spent using social media platforms; Response: <10 min, 10–30 min, 31–60 min, 1–2 h, 2–3 h, 3–5 h, 5–10 h, to more than 10 h	BSI
Fardouly et al. 2020	CS	Investigated differences between preadolescent users and non-users of various social media platforms on mental health.	Australia; 528 (269)	11.19	Self-report daily time spent on social media platform; Response: 0 (0), 1 (<5 min), 2 (5–15 min), 3 (15–30) min, 4 (30 min–1 h), 5 (1–2 h), 6 (2–4 h), 7 (4–6 h), 8 (6–8 h), 9 (8–10 h), 10 (10–12 h or more).	SMFQ
Frison et al. 2016	LS	Examined the relationships between peer victimization on Facebook, depressive symptoms, and life satisfaction.	Belgium; 1621 (51%)	14.76 (12–19)	Self-report daily hours spent on Facebook; Response: 0 (0), 1 (0.5), 2 (0.5–1), 3 (1–1.5), 4 (1.5–2), 5 (2–2.5), 6 (2.5–3), 7 (3–4), 8 (4–5), 9 (>5), 10 (always logged in and available for interaction)	CESD

Table 1. *Cont.*

Study	Design	Main Study Objective	Country; Sample Size (Female)	Age (Years)	Measure of Time Spent on Social Media	Depression Measure
Kelly et al. 2018	CS	Assessed association between social media use and adolescents' depressive symptoms.	UK; 10,904 (50%)	14.30	Self-report daily hours spent on social media; Response: 0, <1, 1–3, 3–5, ≥5	SMFQ
Lemola et al. 2014	CS	Sought a better understand the interplay between sleep, depressive symptoms, and electronic media use at night	Switzerland; 362 (45%)	14.82 (12–17)	Self-report daily duration spent online on Facebook; Response: self-administered open answer	CESD
Ma et al. 2021	LS	Examined how time spent on types of screen use was associated with depressive symptoms.	Sweden; 3556 (51%)	8 grades	Self-report daily hours spent on social media; Response: >2, 2, 1, <1, 0	Question of how often felt depressed
McAllister et al. 2021	CS	Compared associations across specific screen media activities and examined associations with self-harm behaviors.	UK; 4243 (55%)	13.75 (13–15)	Self-report time diary on one weekday and one weekend day from 4:00 am one day to 4:00 am the next day; for each 10 min time slot	SMFQ
Morin-Major et al. 2015	CS	Explored the associations between Facebook and basal levels of cortisol among adolescents.	Canada; 94 (53%)	14.50 (12–17)	Self-report weekly time spent on Facebook; Response (hours): 1 (<1), 2 (2–5), 3 (6–10), 4 (11–15), 5 (16–20), 6 (>21)	CDI
Padilla-Walker et al. 2019	CS	Explored the links between parental media monitoring and adolescents' internalizing symptoms.	USA; 1155 (51%)	10–20	Self-report daily time spent on social media; Response: 1 (none), 2 (less than 30 min), 3 (31–60 min), 4 (1–2 h), 5 (2–3 h), 6 (3–4 h), 7 (5–6 h), 8 (7–8 h), and 9 (≥9 h)	CESD
Pantic et al. 2012	CS	Investigated the relationship between social networking and depression in adolescent.	Serbia; 160 (68%)	18.02	Self-report daily time spent on social networking sites; Response: self-administered open answer	BDI
Sela et al. 2020	CS	Tested the association between family environment and excessive internet use among adolescents.	Israel; 85 (41%)	14.04 (12–16)	Objectively measure time logged in various social medias on the smartphone for 14 days; Response: average time per day spent on social media.	BDI
Shoshani et al. 2021	LS	Examined the influence of the COVID-19 pandemic on children and adolescents' mental health and well-being, and potential risk and protective moderators.	Israel; 1537 (52%)	13.97	Self-report daily hours spent on social media; Response: 0, <1, 1, 2, 3, 4, 5, 6, ≥7.	BSI
Story 2021	CS	Assessed the link between the time spent on social networking sites and depression among 9th and 10th grade high school students.	USA; 85 (56.5%)	14.88 (14–16)	Self-report the number of times and the number of min they spent on SNS daily. Response: sum of the min was divided by the sum of the times	PHQ

Table 1. *Cont.*

Study	Design	Main Study Objective	Country; Sample Size (Female)	Age (Years)	Measure of Time Spent on Social Media	Depression Measure
Tamura et al. 2017	CS	Investigated the relationship between mobile phone use and insomnia and depression in adolescents.	Japan; 295 (41%)	16.20 (15–19)	Self-report daily time spent on social networking sites; Response (min): 0, <30, 30–60, 60–120, ≥120	CESD
Tao et al. 2021	CS	Assessed the relationships among social media use, individual and vicarious social media discrimination, and mental health.	USA; 407 (82%)	16.47 (15–18)	Self-report Total time spent on social media per week; Response: multiple days/week by h/day	CESD
Thorisdottir et al. 2019	CS	Documented the prevalence of social media use and investigate the relationship of both active and passive social media use to anxiety and depressed mood.	Iceland; 10,563 (50%)	14–16	Self-report daily hours on social media; Response: 1 (0) to 8 (≥6)	OSCD
Twenge et al. 2021	CS	Examined associations between different types of screen activities and mental health.	UK; 11,423 (50%)	13.77 (13–15)	Self-report hours spent on social networking or messaging sites on a normal weekday during term time; Response: <0.5, 0.5–0.99, 1–1.99, 2–2.99, 3–4.99, 5–6.99, ≥7	SMFQ
Woods et al. 2016	CS	Examined how social media use related to sleep quality, self-esteem, anxiety and depression.	UK; 467	11–17	Self-report daily hours spent on social media; Response: 1 (<1) to 6 (>6)	HADS
Zielenski et al. 2021	CS	Examined the relationship between Instagram use, social comparison, and depressive symptoms.	USA; 110 (56%)	12–18	Self-report daily hours spent on Instagram; Response:<1 h; 1–2 h; 2–3 h; 3–4 h; 4–5 h; >5 h	CESD

Note: CS, cross-sectional study; LS, longitudinal study; CESD, the Center for Epidemiological Studies-Depression scale; SMFQ, the short version of the Mood and Feelings Questionnaire; BDI, the Beck Depression Inventory; PHQ9, the Patient Health Questionnaire-9; CDI, the Children's Depression Inventory; BSI, the Brief Symptom Inventory; HADS, the Hospital Anxiety and Depression Scale; OSCD, the scale of the Original Symptom Checklist-Depression dimension.

3.3. Subgroup and Sensitivity Analyses

Subgroup analyses show that the association between TSSM and risk of depression was moderated by gender and the measure of depressive symptoms (see Table 2).

Table 2. Moderation analyses for time spent on social media–depression risk association.

Variables	K	OR	95%CI	Z	Heterogeneity Test		
					I^2(%)	Q_w	*p*-Value
Gender, $Q_{b(2)}$ = 40.44 ***							
Boys	4	1.20	1.05–1.37	2.62 *	8.9	3.29	0.349
Girls	4	1.72	1.41–2.09	5.38 ***	66.8	9.03	0.029
Mixed	22	1.67	1.52–1.84	10.27 ***	60.8	53.14	0.001
Age, $Q_{b(2)}$ = 9.28 **							
<14	10	1.54	1.34–1.79	5.85 ***	54.9	19.96	0.018
>14	17	1.61	1.41–1.84	7.10 ***	79	76.11	<0.001
Mixed	3	1.66	1.40–1.97	5.73 ***	0	0.55	0.758
Regions, $Q_{b(3)}$ = 4.13							
Europe	14	1.54	1.33–1.79	5.74 ***	82.8	75.58	<0.001
North America	10	1.68	1.41–1.99	5.88 ***	62.1	23.73	0.005
Asia	4	1.47	1.25–1.73	5.38 ***	0	2.41	0.491
Others	2	1.72	1.41–2.09	4.73 ***	0	0.05	0.820
Measure of Time Spent on Social Media, $Q_{b(1)}$ = 0.23							
Total	26	1.	1.45–1.76	9.39 ***	73.7	95.11	<0.001
Specific	4	1.56	1.01–2.40	1.99	71.6	10.56	0.014
Measure of Depression, $Q_{b(5)}$ = 56.7 ***							
SMFQ	7	1.44	1.26–1.65	5.20 ***	62.3	15.92	0.014
CESD	11	1.77	1.48–2.10	6.39 ***	60	24.98	0.005
BDI	2	1.52	0.96–2.41	1.79	0	0.55	0.458
PHQ9	3	1.55	1.25–1.91	4.04 **	0	1.88	0.391
BSI	3	1.59	1.41–1.80	7.50 ***	36.2	3.14	0.208
Others	4	1.51	1.02–2.24	2.04 *	76.4	12.73	0.005
Sample Sizes, $Q_{b(1)}$ = 0.35							
>1000	13	1.55	1.37–1.76	6.88 ***	83.3	33.5	0.006
<1000	17	1.65	1.42–1.92	6.54 ***	52.3	72.050	<0.001

Note: SMFQ, short version of the Mood and Feelings Questionnaire; CESD, the Center for Epidemiological Studies-Depression scale; BDI, the Beck Depression Inventory, PHQ9, the Patient Health Questionnaire-9; BSI, Brief Symptom Inventory; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

For sensitivity analyses, no single study influenced the result significantly when studies were individually omitted (see Figure S1). There was also no significant change in the results when studies where another effect size was converted to an OR were excluded from analysis (the pooled OR was 1.47, 95%CI: 1.29 to 1.67, $p < 0.001$; $I^2 = 54.4\%$).

3.4. Publication Bias

Begg's test did not show significant publication bias ($p = 0.986$) (see Figure 3), and Egger's linear regression test suggested a mildly significant publication bias ($p = 0.039$). However, no trimming was needed to be performed when the nonparametric trim-and-fill method was used, demonstrating the reliability of the findings (see Figure S2).

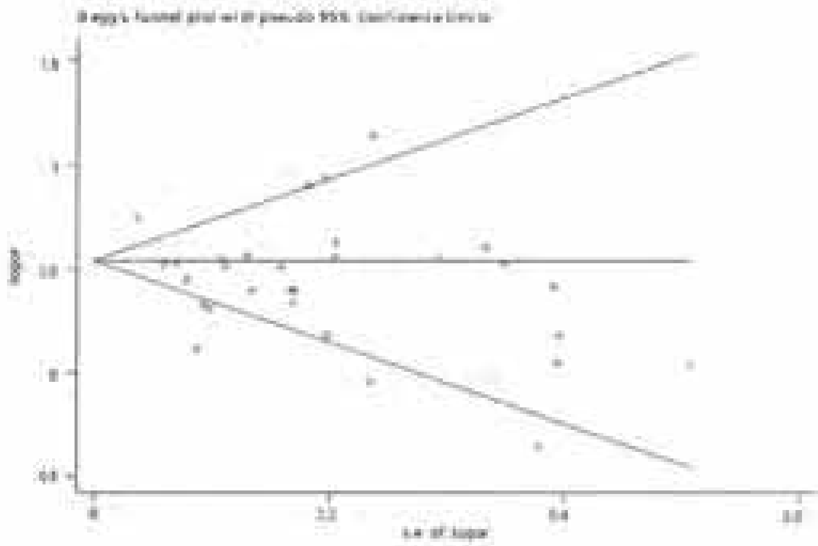


Figure 3. Funnel plot of publication bias.

3.5. *Dose-Response Association between TSSM and Risk of Depression*

Five studies [6,17,19,20,54] (seven reports) were included for the dose-response analysis. The results showed a total linear association between TSSM and risk of depression ($p = 0.888$ for non-linearity, $p < 0.001$ for linearity; see Figure 4) with high heterogeneity between studies ($Q = 70.33$, $p < 0.001$). The risk of depression increased by 13% (OR = 1.13, 95%CI: 1.09 to 1.17, $p < 0.001$) for each hour increase in social media use in adolescents. For samples in which gender was examined separately, there were linear associations between TSSM and depression for both girls and boys ($p = 0.720$ for non-linearity). Specifically, the risk of depression increased by 13% (OR = 1.13, 95%CI: 1.08 to 1.16, $p < 0.001$) for girls and by 9% (OR = 1.09, 95%CI: 1.03 to 1.15, $p = 0.002$) for boys for each hour increase in social media use in adolescents.

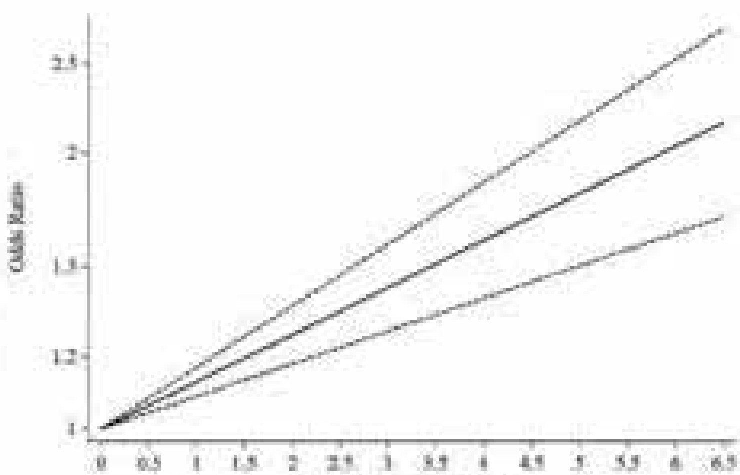


Figure 4. The generalized least squares trend estimated dose-response of time spent on social media and risk of depression in adolescents. Time of social media use was modelled with a restricted cubic spline in a two-stage random-effects dose-response model. The ORs are plotted on the log scale. Dashed lines represent the 95%CIs for the spline model. No social media use served as the referent category.

4. Discussion

The present comprehensive meta-analysis investigated the association between TSSM and the risk of depression in adolescents. Our findings reveal that adolescents with higher daily TSSM had a 59.6% increase in terms of the risk of depression when compared with the reference group. Furthermore, the risk of depression increased by 13% for each hour increase in social media use, and these associations were stronger for adolescent girls than boys; however, boys still demonstrated a significant increase in depression risk. The findings are consistent with the WHO guidelines, which recommend limiting daily screen time for adolescents [65] and which are in agreement with the detrimental effect of high levels of social media use for adolescents suggested by some previous studies [4,5]. Moreover, the linear dose–response analysis of the current study demonstrated that with an increase in hours spent on social media each day, the risk of adolescent depression increased linearly. Therefore, it can be inferred that excessive TSSM may be a strong risk factor for adolescents' depression. Consistently, Twenge et al. [6] suggested that excessive TSSM (>5 h) was associated with a more than 2-fold risk of depression (OR = 2.31, 95%CI: 1.98 to 2.70 for girls; OR = 2.05, 95%CI: 1.59 to 2.64 for boys), after controlling for relevant covariates such as age, family income, ethnicity, and presence of biological father; a similar risk of depression was found in the highest TSSM group (after recalculating using the 0 h/day category as the reference category) in girls (>5 h) in the study conducted by Kelly et al. [19]. Of note, although the included studies in the current meta-analysis have controlled for most relevant covariates (e.g., age, gender, family income, etc.), some other covariates (e.g., physical activity, which was shown to be a protective factor for adolescent' depression) could also influence the results [66,67]. The current findings still need further support by future studies controlling all related covariates.

One notable finding of the current meta-analysis is the significant difference in the pooled estimate between boys and girls. Generally, a significant positive association between TSSM and risk of depression emerges in both girls and boys; however, this pattern is much larger for girls. This finding is consistent with previous studies regarding the association between adolescent social media use and risk of depression [6,68], but it is inconsistent with a longitudinal study examining media exposure (e.g., television, videocassettes, video games, and radio) and risk of depression [69]. In the aforementioned study, the authors found that a lower risk of depression was associated with more total media exposure for teenage girls. Social media use specifically, which was not assessed in this longitudinal study, could underlie these inconsistent results. Social media provides individuals multiple ways of seeking and maintaining social bonds [22,29,45]. Adolescents who fear of missing out social communication hope to continually stay connected with their peers and to stay updated on others' states [70]. Currently, adolescent girls spend much more time on social media than boys [5,6,19,20,32], which may be attributed to the tendency among girls to emphasize close, intimate friendships [28,71]. Therefore, girls are more likely to experience fear of missing out [72] or being harassed [19,20] on social media, which has been associated with risk of depression [28,29]. Thus, it is understandable that previous studies may not have detected gender differences when examining total screen time (including video/computer games, computer/internet use, and television) and risk of depression. Studies stratified by media or screen category and gender are needed to clarify this question.

Higher TSSM was associated with a higher risk of depression in the current meta-analysis across both younger and older adolescence. This seems inconsistent with a previous review [14] in which a significant screen time–depression risk was detected only in younger adolescents (<14 years). One possible reason is that some studies included in the current meta-analysis had a range of both younger and older adolescents [20,27,42,45–47,53]. For example, the study conducted by Woods et al. included participants ranging in age from 11 to 17 years [47]. Some social media platforms have an age limit for creating social media pages (e.g., 13 years of age for Facebook), which may also influence the results. Many studies with no stratification by age precluded us from clarifying the issue yet highlight an important area of future study. The included studies in this meta-analysis also used

multiple different questionnaires measuring depressive symptoms. Interestingly, depression measurement type played a significant moderating role in the association between TSSM and risk of depression. Specifically, the two studies [15,16] in which depression was measured using the BDI demonstrated no significant correlation. The way in which this measure assessed depression may be an important factor to consider. In fact, previous work has asserted that the BDI is a good measure for identifying major depressive disorder [73]; thus, it may not be as accurate when examining a non-clinical population. Similarly, the sample sizes of the two studies were relatively small (n = 85/160), which may be non-comparable and may have caused a misleading correlation. Further studies with a larger sample size are needed to clarify differences in depressive symptoms related to specific measurement scales.

The current meta-analysis comprehensively quantified the dose–response association between TSSM and the risk of depression in adolescents. The large sample size allowed the meta-analysis of dose–response associations between TSSM, ranging from low to high duration, and risk of depression. This study also provided more precise results with smaller confidence intervals than in the previous original studies. International and national guidelines or strategies [74,75] promoting limited screen time for children and adolescents are supported by the current study. Although the implication of the study is not explicit because research is still evolving in this field, our findings reinforce TSSM limitations for adolescents in a gender-specific manner, particularly for girls, noting that the risk of depression increased linearly with an increase in daily hour of social media use. These results are important for adolescents and their parents or other caregivers because they clarify the potential risk of unlimited time on social media, and in turn, prompt them to take steps to promote positive adolescent health and development. Although research concerning links between TSSM and depression in adolescents has given rise to the development of media use policies, particularly regarding smart phone use, most school policies permit their students to use phones during non-instructional times in a school day, such as during recess and lunch [76]. Our findings provide more evidence for school policymakers, as well as national and international public health policymakers, for developing guidelines for appropriate social media use and consumption to reduce depressive risk for adolescents. On the other side, considering the effects of digital technology on the field [77], future study on digital technology innovations is also required for better assisting in "co-care" monitoring adolescents' media use.

There are also important limitations to this study, making it necessary to interpret the findings with caution. First, all included studies were observational, in which the results may be influenced by other potential covariates not yet considered. Hence, we cannot speak to causality in the interpretation of the results. Relatedly, adolescents who had higher depressive symptoms may have recall bias in which they tend to endorse excessive social media use more so than those who had fewer depressive symptoms. As a result, studies in which social media use is measured more objectively are needed in the future. Second, although the search strategies did not restrict language, English databases may lead to the omission of non-English articles as well as non-English terms around social media, which may have an important role in better understanding this association. Although we used a stratified search strategy, there are still various specific social media platforms and social media-related terms, especially those only used in specific countries or regions, that were not included. Third, distinct measurement scales of depressive symptoms and diagnostic criteria for depression could increase variability across included studies. More studies with consistent instruments and diagnostic criteria for depression are needed to support the current findings. Relatedly, understanding the differences in association within a clinical vs. non-clinical population may be important for better understanding for whom TSSM may have a more harmful effect. Most of the participants of eligible studies were collected from Europe and North America, which may limit the generalization of the current findings. Therefore, further investigations from other cultures, particularly focusing on developing countries/regions, are needed to replicate the current findings.

5. Conclusions

Our findings provide evidence that more TSSM is associated with a higher risk of depression in adolescence in a linear dose–response manner, especially for teenage girls. Therefore, prevention efforts targeting a better understanding of the effects of TSSM, particularly for adolescent girls, may be a key component to lessen the risk of depression as social media continues in its global popularity. However, other variables, such as motivation, different platforms, and exposure to social media use may influence this association, making it necessary to interpret the findings with caution. Future research using randomized control studies is required to clarify the causal link between TSSM and depression, as well as the different effects of how adolescents use social media and the environments in which they use it.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijerph19095164/s1. Table S1: The PRISMA checklist of this meta-analysis; Table S2: The risk of bias for included studies; Figure S1: The sensitivity analyses by omitting the included studies one by one; Figure S2: The metatrim analysis by nonparametric trim-and-fill method.

Author Contributions: M.L. contributed to the conception and design, article selection and review, analysis and interpretation of the results, and drafting of the manuscript. K.E.K.-D. contributed to the interpretation of the results, and drafting and revising the manuscript. J.Z. contributed to the interpretation of the results and revision of the manuscript. D.D. contributed to data extraction, interpretation of the results, and revision of the manuscript. J.X. contributed to article selection and review, manuscript preparation, and editing. P.X. contributed to data extraction, manuscript preparation, and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by the Research Foundation of Education Bureau of Hunan Province (Grant No21A0303 to ML). The funding sources had no role in the study design, data collection and analysis, interpretation of the data, preparation and approval of the manuscript, and decision to submit the manuscript for publication.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, by the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kietzmann, J.H.; Hermkens, K.; McCarthy, I.P.; Silvestre, B.S. Social media? Get serious! Understanding the functional building blocks of social media. *Bus. Horiz.* **2011**, *54*, 241–251. [CrossRef]
2. Grajales, I.I.I.F.J.; Sheps, S.; Ho, K.; Novak-Lauscher, H.; Eysenbach, G. Social media: A review and tutorial of applications in medicine and health care. *J. Med. Internet Res.* **2014**, *16*, e13. [CrossRef] [PubMed]
3. Cataldo, I.; Lepri, B.; Neoh, M.J.Y.; Esposito, G. Social Media Usage and Development of Psychiatric Disorders in Childhood and Adolescence: A Review. *Front. Psychiatry* **2021**, *11*, 1332. [CrossRef] [PubMed]
4. Haidt, J.; Allen, N. Digital technology under scrutiny. *Nature* **2020**, *578*, 226–227. [CrossRef] [PubMed]
5. Twenge, J.M.; Haidt, J.; Joiner, T.E.; Campbell, W.K. Underestimating digital media harm. *Nat. Hum. Behav.* **2020**, *4*, 346–348. [CrossRef]
6. Twenge, J.M.; Farley, E. Not all screen time is created equal: Associations with mental health vary by activity and gender. *Soc. Psychiatry Psychiatr. Epidemiol.* **2021**, *56*, 207–217. [CrossRef]
7. Orben, A.; Przybylski, A.K. The association between adolescent well-being and digital technologyuse. *Nat. Hum. Behav.* **2019**, *3*, 173–182. [CrossRef] [PubMed]
8. Pai, P.; Arnott, D.C. User adoption of social networking sites: Eliciting uses and gratifications through a means–end approach. *Comput. Hum. Behav.* **2013**, *29*, 1039–1053. [CrossRef]
9. Deci, E.L.; Ryan, R.M. Self-determination theory: A macrotheory of human motivation, development, and health. *Can. Psychol. Psychol. Can.* **2008**, *49*, 182. [CrossRef]
10. McCallum, C.; McLaren, S. Sense of belonging and depressive symptoms among GLB adolescents. *J. Homosex.* **2010**, *58*, 83–96. [CrossRef]

11. Pillow, D.R.; Malone, G.P.; Hale, W.J. The need to belong and its association with fully satisfying relationships: A tale of two measures. *Personal. Individ. Differ.* **2015**, *74*, 259–264. [CrossRef] [PubMed]
12. Masselink, M.; Van Roekel, E.; Oldehinkel, A. Self-esteem in early adolescence as predictor of depressive symptoms in late adolescence and early adulthood: The mediating role of motivational and social factors. *J. Youth Adolesc.* **2018**, *47*, 932–946. [CrossRef] [PubMed]
13. Nguyen, D.T.; Wright, E.P.; Dedding, C.; Pham, T.T.; Bunders, J. Low self-esteem and its association with anxiety, depression, and suicidal ideation in Vietnamese secondary school students: A cross-sectional study. *Front. Psychiatry* **2019**, *10*, 698. [CrossRef] [PubMed]
14. Liu, M.; Wu, L.; Yao, S. Dose–response association of screen time-based sedentary behaviour in children and adolescents and depression: A meta-analysis of observational studies. *Br. J. Sports Med.* **2016**, *50*, 1252–1258. [CrossRef] [PubMed]
15. Sela, Y.; Zach, M.; Amichay-Hamburger, Y.; Mishali, M.; Omer, H. Family environment and problematic internet use among adolescents: The mediating roles of depression and Fear of Missing Out. *Comput. Hum. Behav.* **2020**, *106*, 106226. [CrossRef]
16. Pantic, I.; Damjanovic, A.; Todorovic, J.; Topalovic, D.; Bojovic-Jovic, D.; Ristic, S.; Pantic, S. Association between online social networking and depression in high school students: Behavioral physiology viewpoint. *Psychiatria Danub.* **2012**, *24*, 90–93.
17. Tamura, H.; Nishida, T.; Tsuji, A.; Sakakibara, H. Association between Excessive Use of Mobile Phone and Insomnia and Depression among Japanese Adolescents. *Int. J. Environ. Res. Public Health* **2017**, *14*, 701. [CrossRef]
18. Banjanin, N.; Banjanin, N.; Dimitrijevic, I.; Pantic, I. Relationship between internet use and depression: Focus on physiological mood oscillations, social networking and online addictive behavior. *Comput. Hum. Behav.* **2015**, *43*, 308–312. [CrossRef]
19. Kelly, Y.; Zilanawala, A.; Booker, C.; Sacker, A. Social Media Use and Adolescent Mental Health: Findings from the UK Millennium Cohort Study. *EClinicalMedicine* **2018**, *6*, 59–68. [CrossRef]
20. Costa BGGd Chaput, J.-P.; Lopes, M.V.V.; Malheiros, L.E.A.; Silva, K.S. Movement behaviors and their association with depressive symptoms in Brazilian adolescents: A cross-sectional study. *J. Sport Health Sci.* **2020**, *11*, 252–259. [CrossRef]
21. Thorisdottir, I.E.; Sigurvinsdottir, R.; Asgeirsdottir, B.B.; Allegrante, J.P.; Sigfusdottir, I.D. Active and Passive Social Media Use and Symptoms of Anxiety and Depressed Mood among Icelandic Adolescents. *Cyberpsychol. Behav. Soc. Netw.* **2019**, *22*, 535–542. [CrossRef] [PubMed]
22. Fardouly, J.; Magson, N.R.; Rapee, R.M.; Johnco, C.J.; Oar, E.L. The use of social media by Australian preadolescents and its links with mental health. *J. Clin. Psychol.* **2020**, *76*, 1304–1326. [CrossRef] [PubMed]
23. Kraut, R.; Patterson, M.; Lundmark, V.; Kiesler, S.; Mukophadhyay, T.; Scherlis, W. Internet paradox: A social technology that reduces social involvement and psychological well-being? *Am. Psychol.* **1998**, *53*, 1017. [CrossRef] [PubMed]
24. Zhang, J. The strain theory of suicide. *J. Pac. Rim Psychol.* **2019**, *13*, e27. [CrossRef]
25. Zhang, J.; Ham, J.M.Y.; Lew, B.; Chistopolskaya, K.; Talib, M.A.; Siau, C.S.; Leung, A.N.M. Depression, anxiety, and stress as a function of psychological strains: Towards an etiological theory of mood disorders and psychopathologies. *J. Affect. Disord.* **2020**, *271*, 279–285. [CrossRef]
26. Liang, L.; Zhou, D.; Yuan, C.; Shao, A.; Bian, Y. Gender differences in the relationship between internet addiction and depression: A cross-lagged study in Chinese adolescents. *Comput. Hum. Behav.* **2016**, *63*, 463–470. [CrossRef]
27. Coyne, S.M.; Rogers, A.A.; Zurcher, J.D.; Stockdale, L.; Booth, M. Does time spent using social media impact mental health?: An eight year longitudinal study. *Comput. Hum. Behav.* **2020**, *104*, 106160. [CrossRef]
28. Murray-Close, D.; Ostrov, J.M.; Crick, N.R. A short-term longitudinal study of growth of relational aggression during middle childhood: Associations with gender, friendship intimacy, and internalizing problems. *Dev. Psychopathol.* **2007**, *19*, 187–203. [CrossRef]
29. Oberst, U.; Wegmann, E.; Stodt, B.; Brand, M.; Chamarro, A. Negative consequences from heavy social networking in adolescents: The mediating role of fear of missing out. *J. Adolesc.* **2017**, *55*, 51–60. [CrossRef]
30. Keles, B.; McCrae, N.; Grealish, A. A systematic review: The influence of social media on depression, anxiety and psychological distress in adolescents. *Int. J. Adolesc. Youth* **2020**, *25*, 79–93. [CrossRef]
31. Zielenski, A.A. Is There a "Happy Filter" on Instagram? The Associations between Instagram Use, Social Comparison, and Depressive Symptoms. Ph.D. Dissertation, Alfred University, Alfred, NY, USA, 2022.
32. Craig, W.; Boniel-Nissim, M.; King, N.; Walsh, S.D.; Boer, M.; Donnelly, P.D.; Harel-Fisch, Y.; Malinowska-Cieślik, M.; de Matos, M.G.; Cosma, A.; et al. Social media use and cyber-bullying: A cross-national analysis of young people in 42 countries. *J. Adolesc. Health* **2020**, *66*, S100–S108. [CrossRef] [PubMed]
33. Piteo, E.M.; Ward, K. Review: Social networking sites and associations with depressive and anxiety symptoms in children and adolescents—A systematic review. *Child. Adolesc. Ment. Health* **2020**, *25*, 201–216. [CrossRef] [PubMed]
34. Sarmiento, I.G.; Olson, C.; Yeo, G.; Chen, Y.A.; Toma, C.L.; Brown, B.B.; Bellmore, A.; Mares, M.L. How Does Social Media Use Relate to Adolescents' Internalizing Symptoms? Conclusions from a Systematic Narrative Review. *Adolesc. Res. Rev.* **2020**, *5*, 381–404. [CrossRef]
35. McCrae, N.; Gettings, S.; Purssell, E. Social Media and Depressive Symptoms in Childhood and Adolescence: A Systematic Review. *Adolesc. Res. Rev.* **2017**, *2*, 315–330. [CrossRef]
36. Ivie, E.J.; Pettitt, A.; Moses, L.J.; Allen, N.B. A meta-analysis of the association between adolescent social media use and depressive symptoms. *J. Affect. Disord.* **2020**, *275*, 165–174. [CrossRef] [PubMed]

37. Moher, D.; Shamseer, L.; Clarke, M.; Ghersi, D.; Liberati, A.; Petticrew, M.; Shekelle, P.; Stewart, L.A. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst. Rev.* **2015**, *4*, 1–9. [CrossRef] [PubMed]
38. Stroup, D.F.; Berlin, J.A.; Morton, S.C.; Olkin, I.; Williamson, G.D.; Rennie, D.; Moher, D.; Becker, B.J.; Sipe, T.A.; Thacker, S.B. Meta-analysis of observational studies in epidemiology: A proposal for reporting. *JAMA* **2000**, *283*, 2008–2012. [CrossRef]
39. Von Elm, E.; Altman, D.G.; Egger, M.; Pocock, S.J.; Gøtzsche, P.C.; Vandenbroucke, J.P. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. *Ann. Intern. Med.* **2007**, *147*, 573–577. [CrossRef]
40. Borenstein, M.; Hedges, L.V.; Higgins, J.P.; Rothstein, H.R. *Introduction to Meta-Analysis*; John Wiley & Sons: Hoboken, NJ, USA, 2021.
41. Hamling, J.; Lee, P.; Weitkunat, R.; Ambühl, M. Facilitating meta-analyses by deriving relative effect and precision estimates for alternative comparisons from a set of estimates presented by exposure level or disease category. *Stat. Med.* **2008**, *27*, 954–970. [CrossRef]
42. Padilla-Walker, L.M.; Stockdale, L.A.; McLean, R.D. Associations between parental media monitoring, media use, and internalizing symptoms during adolescence. *Psychol. Popular Media* **2019**, *9*, 481–492. [CrossRef]
43. Lemola, S.; Perkinson-Gloor, N.; Brand, S.; Dewald-Kaufmann, J.F.; Grob, A. Adolescents' electronic media use at night, sleep disturbance, and depressive symptoms in the smartphone age. *J. Youth Adolesc.* **2014**, *44*, 405–418. [CrossRef] [PubMed]
44. Ellis, W.E.; Dumas, T.M.; Forbes, L.M. Physically Isolated but Socially Connected: Psychological Adjustment and Stress among Adolescents during the Initial COVID-19 Crisis. *Can. J. Behav. Sci. Rev. Can. Des. Sci. Comport.* **2020**, *52*, 177–187. [CrossRef]
45. Dredge, R.; Chen, S. Chinese online gamers versus nongamers: A difference in social media use and associated well-being and relational outcomes? *Psychol. Sch.* **2020**, *57*, 1457–1474. [CrossRef]
46. Morin-Major, J.K.; Marin, M.F.; Durand, N.; Wan, N.; Juster, R.P.; Lupien, S.J. Facebook behaviors associated with diurnal cortisol in adolescents: Is befriending stressful? *Psychoneuroendocrinology* **2016**, *63*, 238–246. [CrossRef] [PubMed]
47. Woods, H.C.; Scott, H. Sleepyteens: Social media use in adolescence is associated with poor sleep quality, anxiety, depression and low self-esteem. *J. Adolesc.* **2016**, *51*, 41–49. [CrossRef]
48. Calandri, E.; Graziano, F.; Rollè, L. Social Media, Depressive Symptoms and Well-Being in Early Adolescence. The Moderating Role of Emotional Self-Efficacy and Gender. *Front. Psychol.* **2021**, *12*, 1871. [CrossRef]
49. McAllister, C.; Hisler, G.C.; Blake, A.B.; Twenge, J.M. Associations between Adolescent Depression and Self-Harm Behaviors and Screen Media Use in a Nationally Representative Time-Diary Study. *JMIR Ment. Health* **2021**, *49*, 1623–1634. [CrossRef]
50. Tao, X. Exposure to Social Media Racial Discrimination and Mental Health among Adolescents of Color. *Vaccines* **2022**, *51*, 30–44. [CrossRef]
51. Story, K. Assessing the Link between Time Spent on Social Networking Sites and Depression among Adolescents. Ph.D. Dissertation, Texas A&M University-Corpus Christi Corpus Christi, Corpus Christi, TX, USA, 2021.
52. Brunborg, G.S.; Andreas, J.B. Increase in time spent on social media is associated with modest increase in depression, conduct problems, and episodic heavy drinking. *J. Adolesc.* **2019**, *74*, 201–209. [CrossRef]
53. Frison, E.; Subrahmanyam, K.; Eggermont, S. The Short-Term Longitudinal and Reciprocal Relations Between Peer Victimization on Facebook and Adolescents' Well-Being. *J. Youth Adolesc.* **2016**, *45*, 1755–1771. [CrossRef]
54. Ma, L.; Evans, B.; Kleppang, A.L.; Hagquist, C. The association between screen time and reported depressive symptoms among adolescents in Sweden. *Fam. Pract.* **2021**, *38*, 773–779. [CrossRef] [PubMed]
55. Shoshani, A.; Kor, A. The mental health effects of the COVID-19 pandemic on children and adolescents: Risk and protective factors. *Psychol. Trauma Theory Res. Pract. Policy* **2021**, 1–9. [CrossRef] [PubMed]
56. Boers, E.; Afzali, M.H.; Newton, N.; Conrod, P. Association of Screen Time and Depression in Adolescence. *JAMA Pediatr.* **2019**, *173*, 853–859. [CrossRef] [PubMed]
57. Fendrich, M.; Weissman, M.M.; Warner, V. Screening for depressive disorder in children and adolescents: Validating the center for epidemiologic studies depression scale for children. *Am. J. Epidemiol.* **1990**, *131*, 538–551. [CrossRef] [PubMed]
58. Angold, A.; Costello, E.J.; Messer, S.C.; Pickles, A.; Winder, F.; Silver, D. Development of a short questionnaire for use in epidemiological studies of depression in children and adolescents. *Int. J. Methods Psychiatr. Res.* **1995**, *5*, 237–249.
59. Beck, A.T.; Steer, R.A.; Ball, R.; Ranieri, W.F. Comparison of Beck Depression Inventories-IA and -II in psychiatric outpatients. *J. Personal. Assess.* **1996**, *67*, 588–597. [CrossRef]
60. Johnson, J.G.; Harris, E.S.; Spitzer, R.L.; Williams, J.B.W. The Patient Health Questionnaire for Adolescents: Validation of an instrument for the assessment of mental disorders among adolescent primary care patients. *J. Adolesc. Health* **2002**, *30*, 196–204. [CrossRef]
61. Timbremont, B.; Braet, C.; Dreessen, L. Assessing depression in youth: Relation between the Children's Depression Inventory and a structured interview. *J. Clin. Child. Adolesc. Psychol.* **2004**, *33*, 149–157. [CrossRef]
62. Derogatis, L.R.; Melisaratos, N. The brief symptom inventory: An introductory report. *Psychol. Med.* **1983**, *13*, 595–605. [CrossRef]
63. Snaith, R.P.; Zigmond, A.S. The hospital anxiety and depression scale. *Br. Med. J.* **1986**, *292*, 344. [CrossRef]
64. Lipman, L.; Covi, R. SCL-90: An outpatient psychiatric rating scale–preliminary report. *Psychopharmacol. Bull.* **1973**, *9*, 13–28.
65. *WHO Guidelines on Physical Activity and Sedentary Behaviour: Web Annex: Evidence Profiles*; World Health Organization: Geneva, Switzerland, 2020.

66. Liu, M.; Zhang, J.; Kamper-Demarco, K.E.; Hu, E.; Yao, S. Associations of moderate-to-vigorous physical activity with psychological problems and suicidality in Chinese high school students: A cross-sectional study. *PeerJ* **2020**, *8*, e8775. [CrossRef] [PubMed]
67. Brown, H.E.; Pearson, N.; Braithwaite, R.E.; Brown, W.J.; Biddle, S. Physical Activity Interventions and Depression in Children and Adolescents. *Sports Med.* **2013**, *43*, 195–206. [CrossRef]
68. Mundy, L.K.; Canterford, L.; Moreno-Betancur, M.; Hoq, M.; Sawyer, S.M.; Allen, N.B.; Patton, G.C. Social networking and symptoms of depression and anxiety in early adolescence. *Depress. Anxiety* **2020**, *38*, 563–570. [CrossRef] [PubMed]
69. Primack, B.A.; Swanier, B.; Georgiopoulos, A.M.; Land, S.R.; Fine, M.J. Association between media use in adolescence and depression in young adulthood: A longitudinal study. *Arch. Gen. Psychiatry* **2009**, *66*, 181–188. [CrossRef] [PubMed]
70. Beyens, I.; Frison, E.; Eggermont, S. "I don't want to miss a thing": Adolescents' fear of missing out and its relationship to adolescents' social needs, Facebook use, and Facebook related stress. *Comput. Hum. Behav.* **2016**, *64*, 1–8. [CrossRef]
71. Rose, A.J.; Rudolph, K.D. A review of sex differences in peer relationship processes: Potential trade-offs for the emotional and behavioral development of girls and boys. *Psychol. Bull.* **2006**, *132*, 98. [CrossRef] [PubMed]
72. Franchina, V.; Vanden Abeele, M.; Van Rooij, A.J.; Lo Coco, G.; De Marez, L. Fear of missing out as a predictor of problematic social media use and phubbing behavior among Flemish adolescents. *Int. J. Environ. Res. Public Health* **2018**, *15*, 2319. [CrossRef]
73. Wilcox, H.; Field, T.; Prodromidis, M.; Scafidi, F. Correlations between the BDI and CES-D in a sample of adolescent mothers. *Adolescence* **1998**, *33*, 565.
74. Bull, F.C.; Al-Ansari, S.S.; Biddle, S.; Borodulin, K.; Buman, M.P.; Cardon, G.; Carty, C.; Chaput, J.-P.; Chastin, S.; Chou, R. World Health Organization 2020 guidelines on physical activity and sedentary behaviour. *Br. J. Sports Med.* **2020**, *54*, 1451–1462. [CrossRef]
75. Tremblay, M.S.; LeBlanc, A.G.; Janssen, I.; Kho, M.E.; Hicks, A.; Murumets, K.; Colley, R.C.; Duggan, M. Canadian sedentary behaviour guidelines for children and youth. *Appl. Physiol. Nutr. Metab.* **2011**, *36*, 59–64. [CrossRef] [PubMed]
76. Tandon, P.S.; Zhou, C.; Hogan, C.M.; Christakis, D.A. Cell Phone Use Policies in US Middle and High Schools. *JAMA Netw. Open* **2020**, *3*, e205183. [CrossRef] [PubMed]
77. Sequeira, L.; Perrotta, S.; LaGrassa, J.; Merikangas, K.; Kreindler, D.; Kundur, D.; Courtney, D.; Szatmari, P.; Battaglia, M.; Strauss, J. Mobile and wearable technology for monitoring depressive symptoms in children and adolescents: A scoping review. *J. Affect. Disord.* **2020**, *265*, 314–324. [CrossRef] [PubMed]

frontiers in Psychology

SYSTEMATIC REVIEW
published: 02 August 2021
doi: 10.3389/fpsyg.2021.641919

The Relationship Between SNS Usage and Disordered Eating Behaviors: A Meta-Analysis

Juan Zhang[1,2], *Yihui Wang*[1,2*], *Qianru Li*[1,2] *and Chenggang Wu*[3]

[1] *Faculty of Education, University of Macau, Macau, China,* [2] *Center for Cognitive and Brain Sciences, University of Macau, Macau, China,* [3] *School of Education, Shanghai International Studies University, Shanghai, China*

Social Networking Sites (SNSs) are common tools with which modern people share their lives and establish social relationships. However, some studies have found SNSs to be associated with eating disorders, although other have identified no connection between the two. To explore the interaction between SNSs and eating disorder behaviors, this study aimed to comprehensively synthesize previous studies using meta-analysis methods. Based on selection criteria, there were 87 effect sizes from 22 studies. After analysis using a three-level random-effects meta-analysis model, a positive correlation between the use of SNSs and irregular eating behaviors was found, $r = 0.09$ (95% CI: 0.06, 0.11; $p < 0.001$). In add potential moderators, body mass
inc 19), survey methods, and sample
so ween SNSs and disordered eating
be sociation between SNSs results
obt ating behaviors ($r = 0.114$; 95%
CI: sults obtained by online surveys
and : −0.102, −0.007; $p < 0.01$).
Univ n SNSs and disordered eating
beha 0.129; $p < 0.001$). Overall, this
meta s associated with an increased
risks study can provide a reference
for th s related to social networks in
the fu

Harms

Keywor rdered eating behaviors

OPEN ACCESS

Edited by:
Shaojing Sun,
Fudan University, China

Reviewed by:
Jake Linardon,
Deakin University, Australia
Xiaomeng Xie,
University of Macau, Macau, China
Wesley Barnhart,
Bowling Green State University, United States

Correspondence:
Yihui Wang
mb94809@um.edu.mo
orcid.org/0000-0002-3221-3013

Specialty section:
This article was submitted to Eating Behavior,
a section of the journal Frontiers in Psychology

Received: *15 December 2020*
Accepted: *03 May 2021*
Published: *02 August 2021*

Citation:
Zhang J, Wang Y, Li Q and Wu C (2021) The Relationship Between SNS Usage and Disordered Eating Behaviors: A Meta-Analysis.
Front. Psychol. 12:641919.
doi: 10.3389/fpsyg.2021.641919

INTR

Eating disorders (EDs) are recognized mental illnesses characterized by irregular eating habits and abnormal concerns about body weight and shapes. Such disorders are chronic, difficult to recover from, prone to relapse and often have serious sequelae (Brownell and Walsh, 2017; Rodgers et al., 2018; Galmiche et al., 2019). Many studies confirmed that EDs make people more vulnerable to psychiatric illnesses such as anxiety and depression, as well as bodily diseases such as diabetes and obesity (Fairburn et al., 2000; Johnson et al., 2001; Schmidt et al., 2016). People with EDs experience a reduced quality of life compared with those suffering from other mental illnesses and physical health conditions (Jenkins et al., 2011a). In severe cases, EDs have been found to be related to

frontiers in Psychology

SYSTEMATIC REVIEW
published: 02 August 2021
doi: 10.3389/fpsyg.2021.641919

The Relationship Between SNS Usage and Disordered Eating Behaviors: A Meta-Analysis

Juan Zhang[1,2], *Yihui Wang*[1,2*], *Qianru Li*[1,2] *and Chenggang Wu*[3]

[1] Faculty of Education, University of Macau, Macau, China, [2] Center for Cognitive and Brain Sciences, University of Macau, Macau, China, [3] School of Education, Shanghai International Studies University, Shanghai, China

Social Networking Sites (SNSs) are common tools with which modern people share their lives and establish social relationships. However, some studies have found SNSs to be associated with eating disorders, although other have identified no connection between the two. To explore the interaction between SNSs and eating disorder behaviors, this study aimed to comprehensively synthesize previous studies using meta-analysis methods. Based on selection criteria, there were 87 effect sizes from 22 studies. After analysis using a three-level random-effects meta-analysis model, a positive correlation between the use of SNSs and irregular eating behaviors was found, $r = 0.09$ (95% CI: 0.06, 0.11; $p < 0.001$). In addition, by analyzing potential moderators, body mass index ($r = -0.032$; 95% CI: −0.058, −0.006; $p = 0.019$), survey methods, and sample sources was discovered could alter the relationship between SNSs and disordered eating behaviors. Specifically, there was a significantly larger association between SNSs results obtained by paper and pencil surveys and disordered eating behaviors ($r = 0.114$; 95% CI: 0.081, 0.147; $p < 0.001$) than that between SNSs results obtained by online surveys and disordered eating behaviors ($r = -0.055$; 95% CI: −0.102, −0.007; $p < 0.01$). University students showed a larger correlation between SNSs and disordered eating behavior than other samples ($r = 0.089$; 95% CI: 0.049, 0.129; $p < 0.001$). Overall, this meta-analysis confirms that the excessive use of SNSs is associated with an increased risks of disordered eating behaviors. It is hoped that this study can provide a reference for the management and intervention of dietary behaviors related to social networks in the future.

Keywords: social networking sites, eating disorder, social media, BMI, disordered eating behaviors

OPEN ACCESS

Edited by:
Shaojing Sun,
Fudan University, China

Reviewed by:
Jake Linardon,
Deakin University, Australia
Xiaomeng Xie,
University of Macau, Macao, China
Wesley Barnhart,
Bowling Green State University, United States

***Correspondence:**
Yihui Wang
yb94850@um.edu.mo
orcid.org/0000-0002-3221-3313

Specialty section:
This article was submitted to Eating Behavior, a section of the journal Frontiers in Psychology

Received: 15 December 2020
Accepted: 03 May 2021
Published: 02 August 2021

Citation:
Zhang J, Wang Y, Li Q and Wu C (2021) The Relationship Between SNS Usage and Disordered Eating Behaviors: A Meta-Analysis. Front. Psychol. 12:641919. doi: 10.3389/fpsyg.2021.641919

INTRODUCTION

Eating disorders (EDs) are recognized mental illnesses characterized by irregular eating habits and abnormal concerns about body weight and shapes. Such disorders are chronic, difficult to recover from, prone to relapse and often have serious sequelae (Brownell and Walsh, 2017; Rodgers et al., 2018; Galmiche et al., 2019). Many studies confirmed that EDs make people more vulnerable to psychiatric illnesses such as anxiety and depression, as well as bodily diseases such as diabetes and obesity (Fairburn et al., 2000; Johnson et al., 2001; Schmidt et al., 2016). People with EDs experience a reduced quality of life compared with those suffering from other mental illnesses and physical health conditions (Jenkins et al., 2011a). In severe cases, EDs have been found to be related to

suicide and other forms of premature mortality (Favaro and Santonastaso, 1997; Bulik et al., 1999). Currently, millions of people suffer from EDs, and the general probability of being affected by ED-related symptoms in one's lifetime is 10% (Schaumberg et al., 2017). Moreover, research has shown that the effects of EDs may reduce family cohesion, increase financial and psychological pressure on family members, and increase psychological risk factors among peers, including the risk of suffering from EDs (Hillege et al., 2006; Keel and Forney, 2013). Disordered eating behaviors represent the core symptoms of EDs. They may consist of implicit attitudes (e.g., eating, weight, and shape concerns) or explicit behaviors (e.g., binge eating, emotional eating, dietary restriction) (Fergus et al., 2019). Whether among young people or adults, irregular eating behaviors are associated with continuous stress, anxiety and other psychological problems, as well as physical health problems such as severe weight fluctuations, which may seriously impair their daily lives (Jenkins et al., 2011b; Neumark-Sztainer et al., 2012; Kärkkäinen et al., 2018). Fueled by these research significances, it is of special interest to examine the risk factors associated with disordered eating behaviors.

Numerous studies have been conducted to investigate the causes of disordered eating behaviors. In this regard, SNSs, as online communication platforms, have become a novel research area of particular interests (Brandtzæg and Heim, 2009). SNSs are among the most popular websites, as identified by statistical websites (Internet Live Stats, 2021). Studies have revealed that, although SNSs bring utility to daily life, their improper use appears to also bring about various problems. Excessive dependence on SNSs can diminish people's satisfaction with their lives, and may increase the likelihood of some people, notably teenagers, feeling depressed and lonely (Spraggins, 2009; Valkenburg and Peter, 2009; Das and Sahoo, 2011). Although several studies have found that concerns about body shape and disordered eating attitudes are positively related to the length of time spent on social media, the results of associated studies have not been consistent (Smith et al., 2013; Mabe et al., 2014; Holland and Tiggemann, 2017). For instance, some discrepancies have been found in the reported relationships between dietary behaviors and different types of SNSs (Kim and Chock, 2015; Blassingame, 2020b).

Although no previous meta-analysis has been conducted in this area, there was one systematic review of 20 articles published before 2016 (Holland and Tiggemann, 2016). This review not only focused on the relationship between SNSs and EDs but also on the relationship between SNSs and body image. It therefore only covered three studies that explicitly discussed the relationship between SNSs and EDs. Although the review concluded that an increase in SNS use was associated with disordered eating behaviors, its results did not indicate the extent to which these two variables were linked.

Since 2018, the related studies have risen sharply in number, and they have begun to explore the interaction between SNSs and EDs in more detail. Some have confirmed the findings of the review mentioned above (Teo and Collinson, 2019; Rodgers et al., 2020), but others have found that the use of SNSs was not directly related to disordered eating behaviors or attitudes (Howard et al., 2017; Cohen et al., 2018; Griffiths et al., 2018a; Blassingame, 2020b). To obtain a better understanding of the relationship between SNS usage and disordered eating behaviors, it is therefore necessary to conduct an up-to-date meta-analysis.

SNS Usage

Development of Social Networking Sites

SNSs are defined as websites or applications located on the Internet that provide individuals with platforms for displaying and sharing their personal lives and interacting with others through provided functions such as comments, likes, and reposts (Perloff, 2014). Compared with traditional media, the emerging Internet-based media provide users with richer information and more diverse communication platforms (Xue and Yu, 2017). SNSs treat users as active participants, not as passive recipients of content gathered from related organizations. They provide users with a great deal of freedom, allowing them to exchange views and develop relationships with each other (Sharma and Verma, 2018). The most popular SNSs in the market, such as Facebook, Twitter, Instagram, and Myspace not only have these qualities, but have also been continuously improved to provide users with a stronger sense of immersion, intimacy, and belonging when they interact with their virtual social circles (Bullas, 2014). As of June 2019, the number of people using SNSs accounted for 72% of the world's population, and this number is increasing every year (Social Networking Fact Sheets, 2019).

Measurement of SNS Usage

Since the focus of this study is on the duration, frequency, and intensity of SNS use, most of the questionnaires it examines are adapted from questionnaires previously used to measure other network activities such as Internet use and Facebook use (Tiggemann et al., 2013; Vanden Abeele et al., 2018). Some questionnaires required participants to answer questions explicitly about the frequency of their use of social software; others asked them to circle a number on the Likert scale that they thought accurately represented this frequency (Griffiths et al., 2018a).

SNS Usage and Disordered Eating Behaviors

Although the way that people participate in SNS activities is completely different from their use of traditional media, research has shown that they might still be as influenced by SNSs as they are by traditional media (Holland and Tiggemann, 2016). For example, studies have revealed that when young females use social media, the pictures they see generally promote slimness as an ideal form of beauty (Tiggemann and Miller, 2010; Fardouly et al., 2015; Brown and Tiggemann, 2016). Previous studies have shown how teenagers' comments and reposts of content containing body stereotypes might induce others to subconsciously approve of this aesthetic and extend its influence in the community (Boyd, 2014). In addition, SNSs may lead people to compare their appearances, body shapes, and affluence. Studies have found that some adolescent girls tend to post their selfies and "outfit-of-the-day" photos on social networking sites, hoping to prove that they are more appealing than their peers

through the comments and compliments they receive (Kaplan and Haenlein, 2010).

Two theoretical frameworks are relevant to explaining the possible relationship between EDs and SNSs. According to sociocultural models, people who interact closely with individuals influence their views about weight and the body (Stice, 1994). These models identify the social and cultural environment work as the most influential and pervasive forces encouraging individuals to promote and pursue ideals of slimness (Rodgers, 2016). In most cases, the figures presented on social media are slim, slender, toned and muscular, which most societies and cultures believe to be the ideal body shape (Carrotte et al., 2017). People struggle to attain this ideal, sometimes under heavy psychological pressure, as not everyone can achieve an idealized body shape, even after heavy exercises and strict control of energy intake (Stice, 1994; Hargreaves and Tiggemann, 2003; Keery et al., 2004; Ata et al., 2007). However, as human often show an instinctive desire to participate in social comparison behaviors automatically and spontaneously (Gilbert and Malone, 1995; Tiggemann, 2012), people tend to show an idealized version of themselves on social media (sometimes an unreal version) to gain recognition from others and from their cultural environment generally. Researchers have found correlations between social interactions in social networks and certain adverse outcomes, not only in terms of personal psychological features (self-esteem, body image, eating habits), but also affecting social relationships (a sense of belonging to the community, sense of happiness, and the ability to get along with others) (Davidson and Cotter, 1991; Obst and Stafurik, 2010). In particular, teenagers who frequently engage in comparisons with peers, or who often request reviews from others on Facebook, are found to be more likely to suffer from eating disorders such as binge eating and overeating (Smith et al., 2013).

Another widely-accepted theoretical basis for the relationship between EDs and SNSs is "self-objectification," by which individuals commoditize their own value, and believes that the value of their personal identity is derived from the use and consumption of their body and its appearance to others (Fredrickson and Roberts, 1997). People may freely post any content that meets social media requirements on SNSs. This freedom has been found to indirectly encourage people (especially young women) to pay greater attention to their appearance than their feelings (Conger and Singg, 2020). Each time users employ the regular functions of SNSs (such as likes and reposts), they may deepen their internalization of a materialized conception of themselves (Chua and Chang, 2016). Research has demonstrated that such a pathological perception of the self is often related to body shame, and may predict an increased likelihood of depression, anxiety, and eating disorders (Przybylski et al., 2013; Beyens et al., 2016; Bell et al., 2018; Ramsey and Horan, 2018). Specifically, Choma et al. (2009), who investigated the self-conception and eating behaviors of female college students, found that body shame contributed to the impact of self-objectification in creating disordered eating. It therefore follows that SNS use may contribute to both self-objectification and body shame, and thus contribute to higher levels of disordered eating behaviors.

Potential Moderators

In view of the sociocultural models, the objectification theory and the varying results summarized above, certain potential moderators should be considered when analyzing the data, to gain a thorough understanding of the relationship between SNS usage and disordered eating behaviors. These moderators are discussed below.

Type of SNS

There are several types of SNS, each with unique designs and functions. Image-centric social media (hereinafter referred to as image-based SNS) are social media whose functions are mainly based on the use of photography and other images. Examples include Instagram, Snapchat and Facebook (Rodgers and Melioli, 2016; Griffiths et al., 2018a). These may be contrasted with social media such as WordPress, whose users are not necessarily expected to display images (Griffiths et al., 2018b). Studies have indicated that, by promoting mutual comparisons among peers, the use of image-based SNSs increases the likelihood that individuals will acquire negative body images (Cohen et al., 2017; Hendrickse et al., 2017; Santarossa and Woodruff, 2017; Marengo et al., 2018). Thus, it may be reasonably inferred that the extent to which a certain SNS is designed based on images may contribute to its users' vulnerability to eating disorders.

Publication Time

Social media have developed rapidly, and new features or entirely new platforms may be appear over a short time (Bowman and Clark-Gordon, 2018). At the same time, the popularity of social media platforms is constantly shifting (Wilksch et al., 2019). A Finnish study found that between 2008 and 2016, the time people spent on SNS use increased and the purpose of their SNS use became more diverse, as did the types of their interactions on SNSs (Koiranen et al., 2019). We therefore infer that publication time is a potential moderator, as growing and widening SNS usage may affect the interaction between SNS and disordered eating behaviors.

Sex

A study pointed out that men and women display different habits in their use of SNSs, and that women tend to post more photos on social platforms, whether selfies, group photos or food photos (Wilksch et al., 2019). People's choices of SNSs may also be related to sex. For example, it was found that teenage girls are more likely to have Instagram and Tumblr accounts than teenage boys (Vannucci and Ohannessian, 2019). Furthermore, although the phenomenon of excessive use of SNSs leading to diet-related problems may be observed in both sexes, a study conducted by Ho et al. (2016) found that females' comparisons with friends and celebrities on SNSs showed a significantly closer relationship to body image dissatisfaction and the drive for slimness than those of males. These respective findings indicate that gender may affect the interaction between SNS use and disordered behaviors.

Region

Given the difference in political system between China and the West, the phenomena of social comparison in Chinese and

Western societies may also differ, as the socialist environment places less emphasis on the individual (Hofstede, 1984; Bandura, 1995; Sedikides et al., 2003). Moreover, different cultures have different traditional aesthetics (Jankowiak et al., 2008). This may explain why, for example, research has shown that compared with Asian adolescents, American adolescents have a deeper internalization of the concept that "beauty is thin" (Leung et al., 2004; Marsh et al., 2007; Klaczynski and Felmban, 2019). Thus, it should be expected that studies on this subject from different regions may well provide different findings.

Age

People of different ages have different life priorities (Erikson, 1994), which might cause adults and adolescents to have different levels of dependence on social media. According to Australian research statistics from 2017, Australian aged between 14 and 17 spend on average half an hour more on social media every day than adults (Australian Psychological Society, 2017). In addition, young people are generally quicker to master new concepts and tools (such as new SNSs) and are more enthusiastic about using SNSs to conduct relationships with others (Kaur et al., 2016; Dhir and Tsai, 2017). Hence, age as a moderating factor in SNS usage may determine individual differences in disordered eating behaviors.

Present Study

To the best of our knowledge, no previous meta-analysis has explored the combined effect of SNS usage and disordered eating behaviors. Given that meta-analysis can quantitatively summarize previous findings (e.g., the correlations between variables) with a large sample size, and can further speculate on the factors that might have affected the relationships between variables by moderator analysis, it is expected that the current meta-analysis will fill the research gap and enhance understanding of the literature on the relationship between disordered eating behaviors and SNS usage. First, this study aimed to collect all the relevant and accessible studies that were conducted before 2020. Second, this research analyzed many potential moderators in order to better explain the inconsistencies between previous research findings. A three-level random-effects meta-analysis model was adopted, which is suitable for obtaining a more accurate evaluation of the overall effect size in a large body of research (Van den Noortgate et al., 2013).

It was hypothesized that individuals experiencing a high intensity or duration of SNS usage would be more likely to exhibit disordered eating behaviors. However, since the results of previous studies did not reach a broad consensus on this subject, further hypothesize could not be on the influence of the moderators on the results of the current study.

METHOD

Literature Search and Study Selection

Related studies were searched and retrieved from five databases (PsychINFO, PubMed, Web of Science, Communication and Mass Media Complete, and ProQuest Dissertations) on July 18th, 2020. The following search keywords for social network usage and disordered eating were chosen: ("social media" OR "social networking sites" OR "SNS" OR "Twitter" OR "Facebook" OR "Weibo" OR "Instagram") AND ("eating" OR "disordered eating" OR "eating disorder"). In addition, a manual search was performed of the reference list in the identified articles to find any other relevant research. It is worth noting that Google Scholar was selected as the fourth resource because it can span multiple disciplines, so that it can be used as a final check to ensure that all articles that meet the current inclusion criteria are captured. Considering the rise of social media, only documents published after 2010 were selected. In fact, no document exceeding this time limit during the search was found. This meta-analysis also aimed to find out the relationship between SNS and disordered eating behaviors in existing studies.

After the initial search, 480 articles were found. The following criteria were applied to screen the 480 articles:

(a) written in English;
(b) published in journals or dissertations;
(c) reported the correlation between social network usage (intensity or frequency) and disordered eating behaviors (r), which had to be a primary goal of the studies.

Only the articles that met the above three criteria were selected.

Coding of Study Features

The following information was retrieved from the selected studies: (1) name of the first author; (2) publication year; (3) age; (4) Body mass index (BMI); (5) percentage of males; (6) percentage of college degree; (7) percentage of white; (8) publication type (dissertation or journal article); (9) region (Western or Eastern); (10) survey methods (paper-and-pencil or online); (11) sample source (university, children and adolescent, clinical or other); (12) SNS type (image-based, non-image-based or general); (13) SNSs usage (duration: time spent on SNSs; frequency: number of times of SNSs usage in a certain period; intensity: integration of SNSs into daily life); (14) type of disordered eating behavior (combined disordered eating behavior, binge eating, driving for thinness, bulimia or dietary restraint); (15) measure of eating (EAT-26, Project Eat III- Eating Behavior Questions, Eating Disorder Inventory, Dutch Eating Behavior Questionnaire, The EDE-Q or others; (16) correlations (r) between SNS and disordered eating behaviors.

To establish internal encoder reliability, two independent coders coded three articles randomly selected from 22 articles. After two rounds of coding, all coders achieved acceptable inter-coder reliability (the Cohen Kappa range of all variables in the coding scheme was 0.85–0.87). They then independently coded the remaining 19 articles and reached an absolute consensus of 95%. The coders resolve any differences through discussion to obtain the final coding result.

Quality Appraisal

Quality assessment was performed by Q.L and Y.W independently. Disagreement about scores was resolved through discussion between the two authors. The quality of the studies included in the meta-analysis was retrieved and adapted

from the previous studies (He et al., 2019, 2020). The adapted tool contained six items: the sampling method, the response rate of the study, the validity of the measurement tool, the source of data, the validity and reliability of examination of SNS usage, and the relationship between SNSs and disordered eating behaviors. Six items were "yes-or-no" questions with a score of 1 for yes and 0 for no. Appraisal scores were obtained by dividing the total score by the total number of items (six), then transforming the fraction into a percentage. In addition, the requirement for the measure of eating pathology relied on those assessments with reliability and validity, including EAT-26, Project EAT III-Eating Behavior Questions, Eating Disorder Inventory, Dutch Eating Behavior Questionnaire and the EDE-Q. This meta-analysis not only included diagnostic samplings, but also other samplings which had not been confirmed as diagnostic patients. Finally, the percentage of appraisal scores in this study was 81.1%, indicating good methodological quality.

To prevent the quality of a selected article from affecting the results, which was analyzed as a categorical moderator in the subsequent analysis (the first category is articles with a quality of 80% and above, and the rest are in the second category). The results show that the quality of the article does not affect the results of this study [$F_{(df1=1,df2=84)} = 3.631, p = 0.060$].

Analysis Plan

As this was a cross-sectional analysis, all data analyses were performed with the R 4.0.0 (R Core Team, 2020) package of metaphor (Viechtbauer, 2010). Raw correlations were converted into Fisher's Zr, since estimate biases would be produced because variance closely relies on the criterion. The converted values were applied in the following analyses. However, Fisher's *Zrs* were back-transformed to correlation coefficients when reporting the results.

The outliers were inspected through the *altimeter* package with the function "meta-outliers" (Lin et al., 2017). A study would be considered as an outlier if the standard residual exceeded 3 (Viechtbauer and Cheung, 2010).

The heterogeneity was assessed by the Cochrane *Q* statistics test, which is a commonly-used index for probing the presence of unexplained heterogeneity (Higgins et al., 2003). Publication bias was examined through Begg's rank correlation test (Egger et al., 1997) and the symmetry of the funnel plot (Duval and Tweedie, 2000).

To avoid dependence problems such as effect sizes, observations, and error terms which are dependent and correlated if they are from the same study, a multilevel meta-analysis was used (Van den Noortgate et al., 2013). The three-level random-effects meta-analytic model was used to discompose variance in different sources. A sample of subjects for each experiment (Level 1), effect sizes within studies (Level 2), and effect sizes changed between studies (Level 3).

RESULTS

Description of Studies Selected

Figure 1 shows a PRISMA flowchart (Hutton et al., 2015) illustrating the procedure for choosing the studies for this meta-analysis. In the searching process, five databases were used to find out appropriate studies. A total of 480 articles were obtained in the initial search. After screening for duplicates, 395 articles were left. In addition, 47 articles were selected through reading abstracts and matching criteria. Eventually, only 22 articles available for full-text review were included in the meta-analysis.

All studies included in this meta-analysis were published between 2010 and 2020. There were 29 independent sample sizes and 87 effect sizes in all 22 studies. A total of 13,301 samples were covered. The sample size of males was 5,031 (37.82% of the total), and 8,270 females. The average age of the sample was between 11.19 and 30.53, and the average BMI varied from 18.92 to 24.69. There were 73 effect sizes obtained from Western and 14 from Eastern. Moreover, 87 effect sizes were reported on the relationship between SNS and disordered eating behaviors (51.72% of the total effect size). Please refer to **Supplementary Data Sheet 1** in Supplementary Material for the coding of this research.

Quality Assessment

Samples, research methods, and data processing for the chosen articles were examined (see **Appendix A** for quality assessment standards). Overall, the included studies' methodological appraisal scores ranged from 66.7 to 100%, which indicated that all studies were of good methodological quality.

Outlier Detection

As **Figure 2** shows, the results of the study conducted by Wilksch et al. (2019) were found to be significantly different from those of the others. Thus, this study was removed from subsequent data analyses as an outliner.

Overall Analysis

From the 22 studies that examined the correlation between SNS and disordered eating behaviors, 87 effect sizes were observed, ranging from −0.35 to 0.45. Significant heterogeneity existed among the effect sizes [$Q_{(df=85)} = 304.27, p < 0.000$] which suggested a need for further moderator analysis to explain heterogeneity. The Forest plot for all samples is presented in **Figure 3**.

Moderator Analysis

Table 1 shows the moderator analysis results for the correlation between SNS and disordered eating behaviors. There were three statistically significant factors: BMI, with $F_{(1,34)} = 6.080$ ($p = 0.019$), Sample source, with $F_{(3,82)} = 2.876$ ($p = 0.041$) and Survey methods, with $F_{(1,84)} = 5.253$ ($p = 0.024$). Although two factors—Region factors, with $F_{(1,84)} = 2.776$ ($p = 0.099$) and Measure of eating, with $F_{(5,80)} = 2.252$ ($p = 0.057$)—were found to approach significance in the moderator analysis, these two factors' effects were >0.05, indicating that neither was significant. The rest of the moderators were found non-significant.

Publication Bias

The Rank Correlation Test for Funnel Plot Asymmetry indicated no publication bias for the correlation between SNS and disordered eating behavior (Kendall's tau = 0.050, $p = 0.530$). The funnel plot is presented in **Figure 4**.

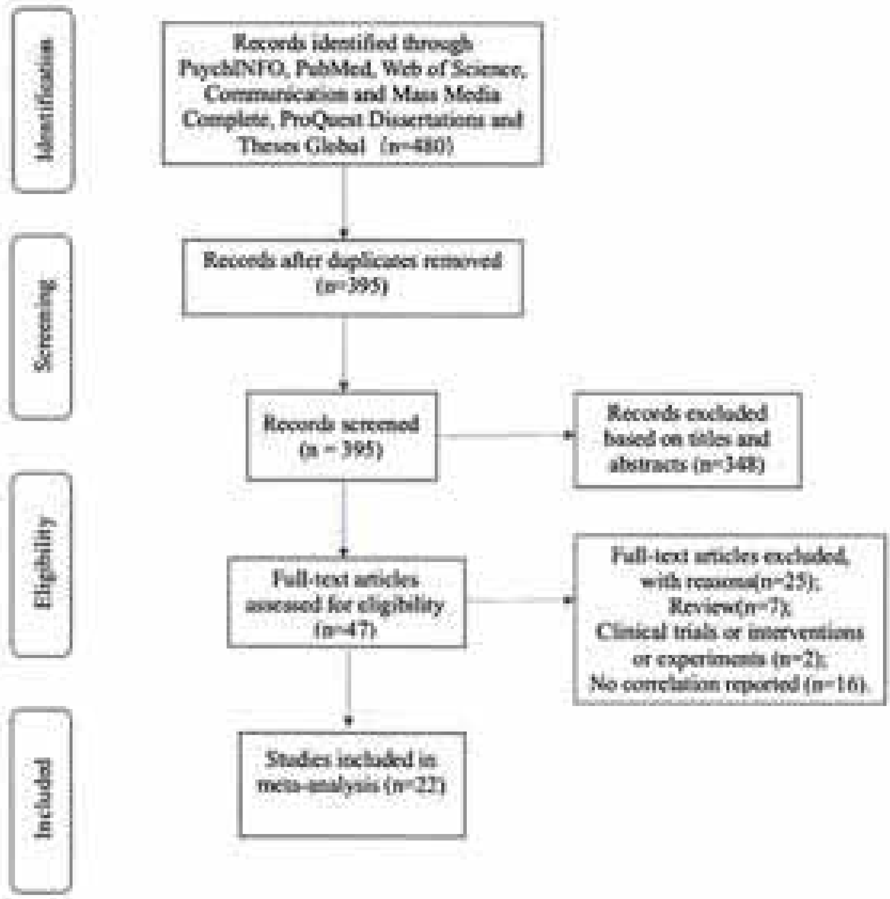


FIGURE 1 | PRISMA flowchart.

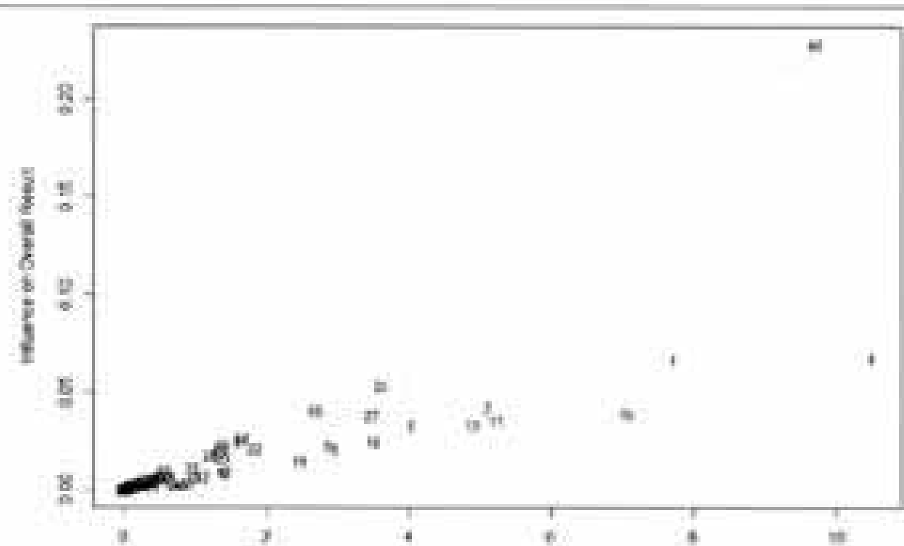


FIGURE 2 | Baujat plot. Notes: each number represents one study included in this research.

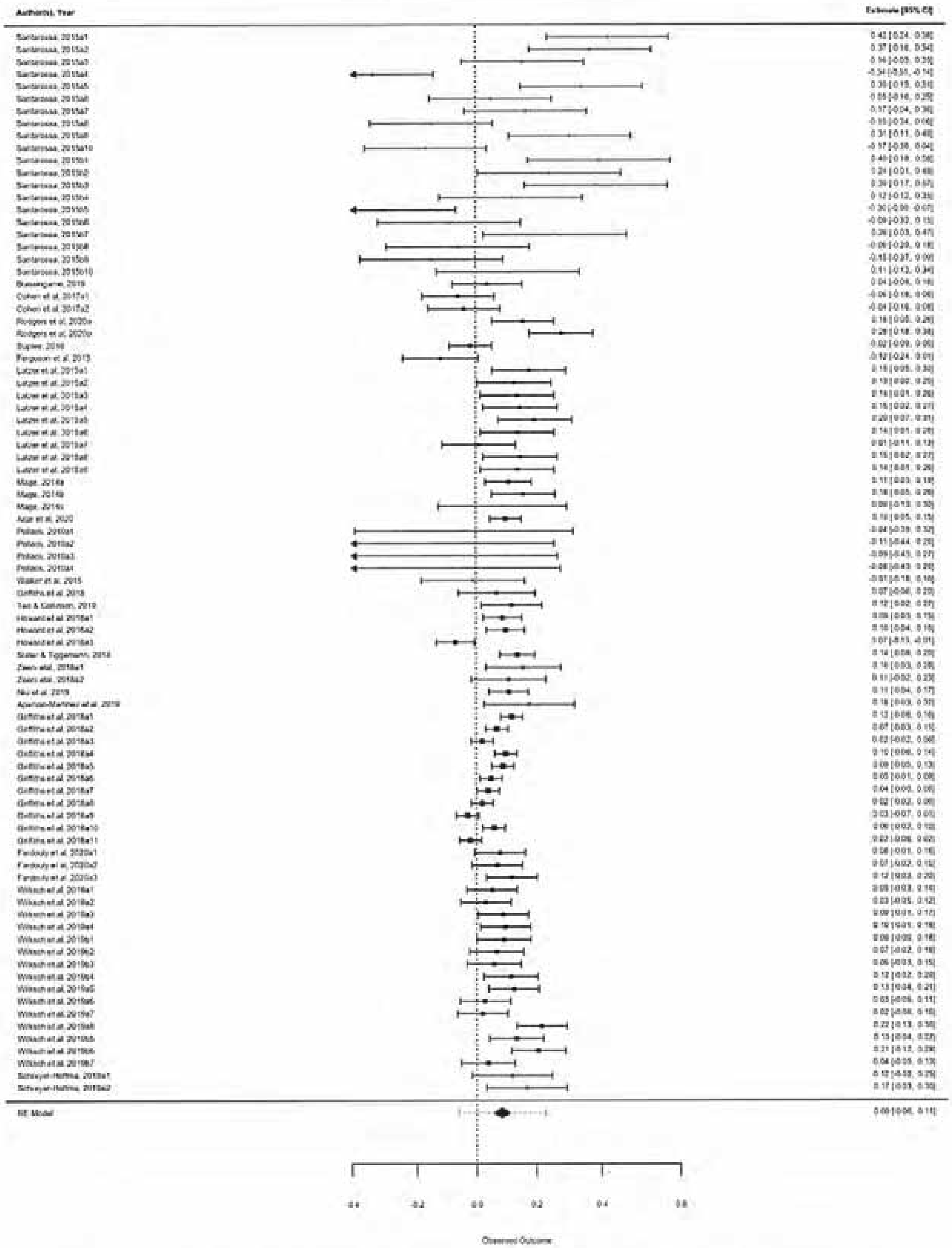


FIGURE 3 | Forest plot for all samples. Notes: The horizontal lines show 95% confidence interval; the diamond represents the point estimate and confidence interval of the pooled effect size.

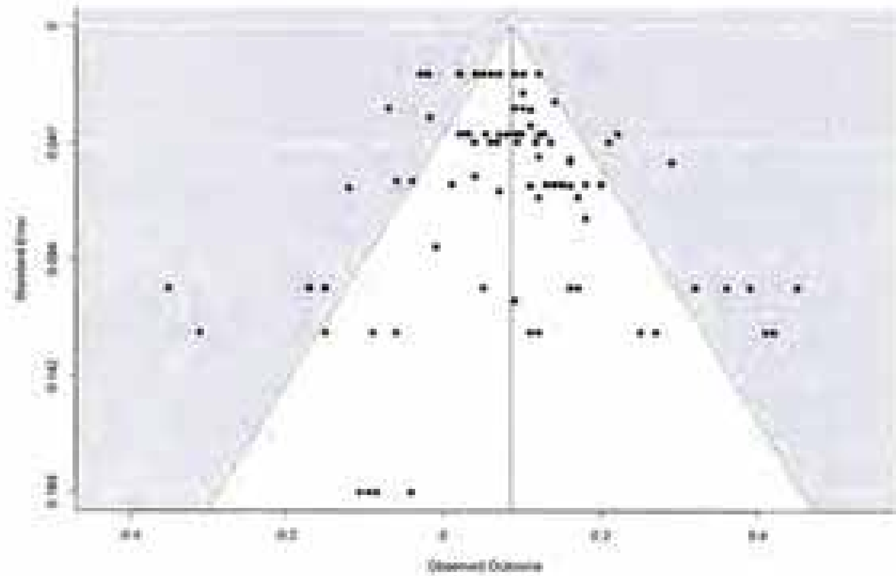


FIGURE 4 | Funnel plot. Notes: Black circles are the studies included in the meta-analysis.

DISCUSSION

The current study was intended to expand upon previous work by adopting a three-level meta-analysis model to analyze the association between SNSs and disordered eating behaviors. Analysis in this study revealed a weak but significant positive correlation between the use of SNSs and disordered eating behaviors, in line with the results of several previous studies (Mabe et al., 2014; Latzer et al., 2015; Santarossa, 2015; Aparicio-Martinez et al., 2019; Niu et al., 2019; Teo and Collinson, 2019; Blassingame, 2020b; Rodgers et al., 2020). Considering that the focus of analysis in this study was the influence of frequency and duration of SNS usage on eating disorders, combined with the social-cultural model and self-objectification theory, the high frequency and long-term use of SNSs might indeed lead people to participate regularly in social comparisons (Ho et al., 2016). Excessive immersion in the appearance comparisons in the surrounding environment may make it more difficult for people to generate a positive body image, and may also lead people to suffer from stress relating to appearance and body shape, leading in turn to various mental illnesses, including disordered eating behaviors (Tylka and Subich, 2010).

Moderation models help to understand if other variables explain the strength of relationship between two variables. This study therefore assessed several potential moderators, hoping to gain a clearer understanding of the inconsistency of findings in the literature as to whether and to what extent SNS usage is related to disordered eating behaviors. Moderator analyses showed that the sample source, survey method, and mean BMI of the sample were the significant moderators, which may explain the individual differences of the previous findings.

As for the sample source, university students were the main contributors to the discrepancies in disordered eating behaviors associated with SNS usage, compared with children, adolescents, clinical samples, and other samples. College students are the main contributors to studies investigating social media use (Zhang and Leung, 2015). For university students, social media are not only communication tools, but also form an important part of their daily routines (Madge et al., 2009). Regular use of SNSs may amplify the impact of SNSs on university students' lives, either positively or negatively (Gosling and Mason, 2015). At the same time, research has shown that some college students often use certain SNSs (such as Snapchat and Instagram) to make or receive appearance-related comparisons or comments (Verduyn et al., 2015). In other words, the purpose and of college students' use of social networks seems to vary the impact of the social networks has on their body image, which may further lead to further differences in their disordered eating habits. Therefore, when investigating the social media and disordered eating behaviors of college students, the interactions between SNSs usage and disordered eating behaviors are not in close accordance.

The survey methods are likely to also adjust the association between disordered eating behaviors and SNSs usages between different subgroups, because sometimes measuring the same variable through different methods (such as the pen-and-paper and online methods mentioned in this article) may yield different results (Moessner et al., 2015). Many tools for measuring eating disorder behaviors, including some commonly used tools, such as EAT by Garner and Garfinkel (1979), and BULIT-R by Thelen et al. (1991), have no dedicated online versions. When comparing the results of these questionnaire surveys in the laboratory with those on the Internet, some studies found that the two are highly similar, but some studies have obtained the opposite results

TABLE 1 | Moderator analyses for studies reporting the correlation between SNS and disordered eating behaviors.

Moderator variables	#Studies	#ES	β_0 (95% CI)	ESr	β_1 (95% CI)	F(df1,df2)	Level 2 variance	Level 3 variance
Publication year	27	87	−11.707 (−32.051; 8.638)		0.006 (−0.004; 0.016)	1.329 (1, 84)	0.005***	0.001
Age	18	36	0.126 (−0.023; 0.276)		−0.003 (−0.011; 0.005)	0.523 (1, 34)	0.002**	0.002
BMI	13	37	0.746 (0.199; 1.292)**		−0.032(−0.058; −0.006)*	6.080 (1, 34)*	0.001	0.002
Percent of male	11	39	0.089 (0.057; 0.121)		−0.007 (−0.058;0.045)	0.065 (1, 84)	0.005***	0.001
Percent of college	14	37	0.107 (0.071; 0.144)		0.001 (−0.000; 0.003)	2.194 (1, 32)	0.007***	0.000
Percentage of white	14	23	0.049 (−0.049; 0.147)		0.008(−0.136; 0.152)	0.014 (1, 23)	0.002**	0.003
Publication type						0.113(1, 84)	0.005***	0.001
Journal	21	59	0.089 (0.057; 0.121)	0.088				
Thesis	6	28	−0.012 (−0.085; 0.060)	−0.0120	−0.101			
Region						2.776 (1, 84)*	0.005***	0.000
Western	22	73	0.077 (0.050; 0.104)***	0.0768***				
Eastern	5	14	0.052 (−0.010; 0.114)	0.0520	−0.025			
Survey methods						5.253 (1, 84)*	0.004***	0.000
Paper-and-pencil	16	43	0.114 (0.081; 0.147)***	0.114***				
Online	11	44	−0.055 (−0.102; −0.007)*	−0.055*	−0.169			
Sample source						2.876 (3, 82)*	0.005***	0.000
University sample	12	35	0.089 (0.049; 0.129)***	0.089***				
Children and adolescent sample	7	29	0.035 (−0.023; 0.092)	0.035	−0.054			
Clinical sample	1	4	−0.171 (−0.379; 0.038)	−0.169	−0.26			
Other sample	7	19	−0.039 (−0.099; 0.021)	−0.039	−0.128			
SNS use measure						0.105 (3, 82)	0.005***	0.001
Duration	13	51	0.084 (0.040; 0.127)	0.084				
Frequency	10	32	0.010 (−0.052; 072)	0.010	0.074			
Intensity	2	2	−0.014 (−0.171; 0.143)	−0.014	−0.098			
Mixed	2	2	−0.024 (−0.170; 0.122)	−0.024	0.108			
SNS type							0.005	0.002
Image-based	10	34	0.070 (0.027; 0.112)*	0.070				
Non image-based	13	31	−0.234 (−0.334; −0.134)***	−0.223	−0.304			
General	14	22	0.044(−0.021,108)	0.044	−0.026			
Type of disordered eating						0.712(4, 82)	0.006***	0.000
Combined disordered eating behavior	22	54	0.077 (0.047; 0.108)***	0.077				
Binge eating	4	18	0.028 (−0.033; 0.090)	0.028	−0.049			
Driving for thinness	3	5	0.005 (−0.086; 0.097)	0.005	−0.072			
Bulimia	2	4	0.025 (−0.080; 0.131)	0.025	−0.052			
Dietary restraint	5	6	0.069 (−0.025; 0.164)	0.069	−0.008			

(Continued)

TABLE 1 | Continued

Moderator variables	#Studies	#ES	β_0 (95% CI)	ESr	β_1 (95% CI)	$F(df1,df2)$	Level 2 variance	Level 3 variance
Measure of eating						2.252(5, 80)^	0.005***	0.000
EAT-26	13	26	0.109 (0.068; 0.150)***	0.109				
Project Eat III–Eating behavior questions	2	10	−0.005 (−0.099; 0.089)	−0.005	−0.114			
Eating disorder Inventory	4	11	−0.018 (−0.091; 0.055)	−0.018	−0.127			
Dutch eating behavior questionnaire	3	3	0.073 (−0.033; 0.180)	0.073	−0.036			
The EDE-Q	9	28	−0.058 (−0.109; −0.008)*	−0.058	−0.167			
Others	3	10	0.007 (−0.064; 0.078)	0.007	−0.102			

#Studies, Number of studies; #ES, Number of effect sizes; CI, Confidence interval; Level 2 variance, Variance in effect sizes within studies; Level 3 variance, Variance in effect sizes between studies.
^$p < 0.1$.
**$p < 0.05$.*
***$p < 0.01$.*
****$p < 0.001$.*

(Joinson, 1999; Rammstedt et al., 2004). Therefore, although online surveys can collect data more conveniently and quickly from various channels if the survey tool used is not a dedicated online version, the results obtained may be different from the data collected face-to-face.

As far as BMI is concerned, the results indicate that an increase in the average BMI of the sample is usually accompanied by a decrease in the tendency for people to suffer from eating disorders associated with SNS use. BMI has been shown to have a significant impact on an individual's eating behaviors (Burnette et al., 2018). Unlike the positive relationship between BMI and disordered eating behaviors claimed by most research, this result indicates that a higher level of BMI may decrease the possibility of people having irregular eating behaviors after using of SNSs for a long time or in high intensity (Goldschmidt et al., 2008). Since there are no existing research results that allow us to explain this mechanism, we can only deduce that people with higher BMI may avoid SNS-related social behaviors, and that they may thus be exposed to fewer negative impacts from SNS that may lead to disordered eating behaviors. Since the figures of people with a lower BMI are more in line with the ideal body shape in most cultural environments, they might be more inclined to participate in the body comparisons on SNSs and further internalize the ideal slimness value, which may result in a negative view of their body and lead in turn to disordered eating behaviors and attitudes. Similarly, Yao et al. (2021) argue that people with a low BMI are more likely to demonstrate restrained eating behaviors because they lack confidence in their body. It is therefore reasonable to infer that people with a lower BMI could be more vulnerable to disordered eating behaviors when participating in activities related to SNSs.

In conclusion, the present meta-analysis has offered a quantitative synthesis of the current state of knowledge on the relationship between SNS usage and disordered eating behaviors. Based on a three-level meta-analytic model and moderator analysis, the research has demonstrated that SNS use is significantly linked to disordered eating behaviors and attitudes, which might be altered by the sample source, survey method, and mean BMI of the sample. According to sociocultural theory and self-objectification theory, individuals who use SNSs frequently and intensively seems to be more likely to internalize the ideal value of slimness of their social and cultural environment through information in SNSs, and to take part in social comparisons related to appearance. On the other hand, SNS usage might encourage individuals to connect their values with their body shapes. Both these inferences suggest that SNS usage is likely to lead to body dissatisfaction and may indeed play a causal role in the development of disordered eating behaviors.

LIMITATIONS AND FUTURE RESEARCH DIRECTIONS

Some limitations should be considered when interpreting the current research. First, language was set to English when filtering the research. In addition, some gray literature (such as ongoing research) would not appear in the general search process. This

may have caused some literature to be lost from analysis of this study. Therefore, future meta-analyses should include richer studies to conduct a more comprehensive and thorough analysis of related issues. Second, in addition to the proposed moderator, there might be other factors that affect the consistency of the research results, such as sexual orientation and the ways of using SNSs (Ryding and Kuss, 2019; He et al., 2020). As there were few relevant studies, the influence of these factors on the relationship between SNS and eating disorders could not be analyzed, which awaits future research to fill these gaps. Third, because not every study contains the potential moderators that this meta-analysis set, this mean that the moderator analysis results are not applicable to all the studies in the relevant field. It therefore remains an important direction for future research to include more factors when identifying the situations and mechanisms concerning whether and how SNS usage is related to individuals' body dissatisfaction.

Future research needs to explore the other popular and current forms of SNSs (i.e., Twitter, Instagram, and Pinterest) owing to the rapid development of SNS platforms (Duggan et al., 2015). Moreover, specific ways (active/passive) of using SNSs should also be considered when conducting future research, to enhance understandings of the mechanisms and situations affecting whether and how SNS usage is associated with disordered eating.

CONCLUSION AND IMPLICATIONS

The current meta-analysis revealed a small, positive correlation between frequent and intensive use of SNSs and disordered eating behaviors. In addition, the BMI of the sample, the source of the sample, and the survey method (paper-and-pencil or online) were identified as the moderators that may explain the inconsistent findings between SNSs usage and disordered eating.

The findings from this meta-analysis have several clinical implications. First, this study found a positive correlation between the use of SNS and disordered eating behaviors. Clinicians may therefore consider evaluating the influence of SNS use on patients' irregular eating behaviors during the intervention process for disordered eating behaviors, and intervene with a view to controlling the length and frequency of SNS use.

The current meta-analysis also draws attention to the importance of the proper usage of SNSs in preventing disordered eating behaviors. According to the sociocultural theory and objectification theories which may explain the underlying principles of SNS usage and their positive associations with disordered eating behaviors, in addition to the need to control the frequency and duration of use of SNSs, the overall aesthetic orientation of these media and the comments made in them by the others are what affect the consequences of using SNSs. Frequent social comparisons on social media may aggravate the conflicts between the glamorous social images that people see displayed on their homepages and their perceptions of themselves. The findings of this meta-analysis may therefore have several media-related implications. Relevant institutions are advised to regularly organize some positive campaigns on SNSs to encourage people to pay attention to personal characteristics other than mere appearance. The promotion of the positive use of SNSs, which has been connected to fewer negative outcomes such as eating disorders and more positive outcomes such as the formation of social bonds, is also recommended (Verduyn et al., 2017). It is hoped that the current research can draw attention to the need to create a positive and healthy network environment for netizens in which SNSs facilitate their communication and help to establish their social relationships.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

YW and JZ contributed to the research design. YW and QL collected relevant articles, completed the coding, and drafted the manuscript. YW analyzed the data. JZ and CW carefully revised the manuscript. All authors read and approved the final version of the manuscript.

FUNDING

This work was supported by the University of Macau [MYRG2017-00217-FED] in Macau.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.641919/full#supplementary-material

REFERENCES

Acar, M., Akgul, S., Kizilkan, M. P., Tuzun, Z., Derman, O., and Kanbur, N. (2020). Eating attitudes and physical appearance comparison with others in daily life versus on social media in adolescents. *J. Adolesc. Health* 66, S59–S60.

Aparicio-Martinez, P., Perea-Moreno, A. J., Martinez-Jimenez, M. P., Redel-Macias, M. D., Pagliari, C., and Vaquero-Abellan, M. (2019). Social media, thin-ideal, body dissatisfaction and disordered eating attitudes: an exploratory analysis. *Int. J. Environ. Res. Public Health* 16:4177. doi: 10.3390/ijerph16214177

Ata, R. N., Ludden, A. B., and Lally, M. M. (2007). The effects of gender and family, friend, and media influences on eating behaviors and body image during adolescence. *J. Youth Adolesc.* 36, 1024–1037. doi: 10.1007/s10964-006-9159-x

Australian Psychological Society (2017). *Digital Me: A Survey Exploring the Effect of Social Media and Digital Technology on Australians' Wellbeing.* Melbourne, VIC: Australian Psychological Society.

Bandura, A. (1995). "Exercise of personal and collective efficacy in changing societies," in *Self-Efficacy in Changing Societies* (New York, NY: Cambridge University Press), 1–45. doi: 10.1017/CBO9780511527692.003

Bell, B. T., Cassarly, J. A., and Dunbar, L. (2018). Selfie-objectification: Self-objectification and positive feedback ("likes") are associated with frequency of posting sexually objectifying self-images on social media. *Body Image* 26, 83–89. doi: 10.1016/j.bodyim.2018.06.005

Beyens, I., Frison, E., and Eggermont, S. (2016). "I don't want to miss a thing": Adolescents' fear of missing out and its relationship to adolescents' social needs, Facebook use, and Facebook related stress. *Comput. Hum. Behav.* 64, 1–8. doi: 10.1016/j.chb.2016.05.083

Blassingame, J. C. I., II. (2020a). *Yassforthebody: Social media, body dissatisfaction, and disordered eating among gay men (Order No. 22617574).* Available from ProQuest Dissertations & Theses Global (2313355484). Retrieved from: https://libezproxy.um.edu.mo/login?url=https://www.proquest.com/dissertations-theses/yassforthebody-social-media-body-dissatisfaction/docview/2313355484/se-2?accountid=12206

Blassingame, J. C. III. (2020b). *#Yassforthebody: Social Media, Body Dissatisfaction, and Disordered Eating among Gay Men.* (Ph.D.). Fielding Graduate University, Ann Arbor. Retrieved from: https://search.proquest.com/docview/2313355484?accountid=12206 (accessed May 30, 2020). ProQuest Dissertations & Theses Global database. (22617574)

Bowman, N. D., and Clark-Gordon, C. V. (2018). "A brief history of social media: evolution and impact," in *Social Media: Practices, Uses Global Impact*, ed T. C. Ahern (New York, NY: Nova Science Publishers), 1–16.

Boyd, D. (2014). *It's Complicated: The Social Lives of Networked teens.* New Haven, CT: Yale University Press.

Brandtzæg, P. B., and Heim, J. (2009). "Why people use social networking sites," in *Paper Presented at the International Conference on Online Communities and Social Computing* (Oslo). doi: 10.1007/978-3-642-02774-1_16

Brown, Z., and Tiggemann, M. (2016). Attractive celebrity and peer images on Instagram: effect on women's mood and body image. *Body Image.* 19, 37–43. doi: 10.1016/j.bodyim.2016.08.007

Brownell, K. D., and Walsh, B. T. (2017). *Eating Disorders and Obesity: A Comprehensive handbook.* New York, NY: Guilford Publications.

Bulik, C. M., Sullivan, P., and Joyce, P. R. (1999). Temperament, character and suicide attempts in anorexia nervosa, bulimia nervosa and major depression. *Acta Psychiatr. Scand.* 100, 27–32. doi: 10.1111/j.1600-0447.1999.tb10910.x

Bullas, J. (2014). *22 Social Media Facts and Statistics You Should Know in 2014.* Retrieved from: http://www.jeffbullas.com/2014/01/17/20-social-media-facts-and-statistics-you-should-know-in-2014/ (accessed October 17, 2015).

Burnette, C. B., Simpson, C. C., and Mazzeo, S. E. (2018). Relation of BMI and weight suppression to eating pathology in undergraduates. *Eat. Behav.* 30, 16–21. doi: 10.1016/j.eatbeh.2018.05.003

Carrotte, E. R., Prichard, I., and Lim, M. S. C. (2017). "Fitspiration" on social media: a content analysis of gendered images. *J. Med. Internet Res.* 19:e95. doi: 10.2196/jmir.6368

Choma, B. L., Shove, C., Busseri, M. A., Sadava, S. W., and Hosker, A. (2009). Assessing the role of body image coping strategies as mediators or moderators of the links between self-objectification, body shame, and well-being. *Sex Roles* 61:699. doi: 10.1007/s11199-009-9666-9

Chua, T. H. H., and Chang, L. C. (2016). Follow me and like my beautiful selfies: Singapore teenage girls' engagement in self-presentation and peer comparison on social media. *Comput. Hum. Behav.* 55, 190–197. doi: 10.1016/j.chb.2015.09.011

Cohen, R., Newton-John, T., and Slater, A. (2017). The relationship between Facebook and Instagram appearance-focused activities and body image concerns in young women. *Body Image.* 23, 183–187. doi: 10.1016/j.bodyim.2017.10.002

Cohen, R., Newton-John, T., and Slater, A. (2018). 'Selfie'-objectification: the role of selfies in self-objectification and disordered eating in young women. *Comput. Hum. Behav.* 79, 68–74. doi: 10.1016/j.chb.2017.10.027

Conger, J., and Singg, S. (2020). Should you "do it for the 'gram?" A review of the psychological and sociocultural effects of social media. *J. Clin. Rev. Case Rep.* 5:9. doi: 10.33140/jcrc.05.01.09

Das, B., and Sahoo, J. S. (2011). Social networking sites–a critical analysis of its impact on personal and social life. *Int. J. Business Soc. Sci.* 2, 222–228. Retrieved from https://ijbssnet.com/journal/index/561

Davidson, W. B., and Cotter, P. R. (1991). The relationship between sense of community and subjective well-being: a first look. *J. Commun. Psychol.* 19, 246–253. doi: 10.1002/1520-6629(199107)19:3<246::AID-JCOP2290190308>3.0.CO;2-L

Dhir, A., and Tsai, C.-C. (2017). Understanding the relationship between intensity and gratifications of Facebook use among adolescents and young adults. *Telematics Inform.* 34, 350–364. doi: 10.1016/j.tele.2016.08.017

Duggan, M., Ellison, N. B., Lampe, C., Lenhart, A., and Madden, M. (2015). Social media update 2014. *Pew Res. Center* 19, 1–2. Retrieved from https://www.pewresearch.org/internet/2015/01/09/social-media-update-2014/

Duval, S., and Tweedie, R. (2000). Trim and fill: a simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 56, 455–463. doi: 10.1111/j.0006-341X.2000.00455.x

Egger, M., Smith, G. D., Schneider, M., and Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ* 315, 629–634. doi: 10.1136/bmj.315.7109.629

Erikson, E. H. (1994). *Identity and the Life Cycle.* New York, NY: WW Norton & Company.

Fairburn, C. G., Cooper, Z., Doll, H. A., Norman, P., and O'Connor, M. (2000). The natural course of bulimia nervosa and binge eating disorder in young women. *Arch. Gen. Psychiatry* 57, 659–665. doi: 10.1001/archpsyc.57.7.659

Fardouly, J., Diedrichs, P. C., Vartanian, L. R., and Halliwell, E. (2015). Social comparisons on social media: The impact of Facebook on young women's body image concerns and mood. *Body Image* 13, 38–45. doi: 10.1016/j.bodyim.2014.12.002

Fardouly, J., Magson, N. R., Rapee, R. M., Johnco, C. J., and Oar, E. L. (2020). The use of social media by Australian preadolescents and its links with mental health. *J. Clin. Psychol.* 76, 1304–1326. doi: 10.1002/jclp.22936

Favaro, A., and Santonastaso, P. (1997). Suicidality in eating disorders: clinical and psychological correlates. *Acta Psychiatr. Scand.* 95, 508–514. doi: 10.1111/j.1600-0447.1997.tb10139.x

Fergus, K. B., Copp, H. L., Tabler, J. L., and Nagata, J. M. (2019). Eating disorders and disordered eating behaviors among women: associations with sexual risk. *Int. J. Eating Disord.* 52, 1310–1315. doi: 10.1002/eat.23132

Ferguson, C. J., Muñoz, M. E., Garza, A., and Galindo, M. (2013). Concurrent and prospective analyses of peer, television and social media influences on body dissatisfaction, eating disorder symptoms and life satisfaction in adolescent girls. *J. Youth Adolesc.* 43, 1–14. doi: 10.1007/s10964-012-9898-9

Fredrickson, B. L., and Roberts, T. A. (1997). Objectification theory: toward understanding women's lived experiences and mental health risks. *Psychol. Women Quart.* 21, 173–206. doi: 10.1111/j.1471-6402.1997.tb00108.x

Galmiche, M., Déchelotte, P., Lambert, G., and Tavolacci, M. P. (2019). Prevalence of eating disorders over the 2000–2018 period: a systematic literature review. *Am. J. Clin. Nutr.* 109, 1402–1413. doi: 10.1093/ajcn/nqy342

Garner, D. M., and Garfinkel, P. E. (1979). The eating attitudes test: an index of the symptoms of anorexia nervosa. *Psychol. Med.* 9, 273–279. doi: 10.1017/S0033291700030762

Gilbert, D. T., and Malone, P. S. (1995). The correspondence bias. *Psychol. Bull.* 117:21. doi: 10.1037/0033-2909.117.1.21

Goldschmidt, A. B., Aspen, V. P., Sinton, M. M., Tanofsky-Kraff, M., and Wilfley, D. E. (2008). Disordered eating attitudes and behaviors in overweight youth. *Obesity* 16, 257–264. doi: 10.1038/oby.2007.48

Gosling, S. D., and Mason, W. (2015). Internet research in psychology. *Annu. Rev. Psychol.* 66, 877–902. doi: 10.1146/annurev-psych-010814-015321

Griffiths, S., Castle, D., Cunningham, M., Murray, S. B., Bastian, B., and Barlow, F. K. (2018a). How does exposure to thinspiration and fitspiration relate to symptom severity among individuals with eating disorders? Evaluation of a proposed model. *Body Image* 27, 187–195. doi: 10.1016/j.bodyim.2018.10.002

Griffiths, S., Murray, S. B., Krug, I., and McLean, S. A. (2018b). The contribution of social media to body dissatisfaction, eating disorder symptoms, and anabolic steroid use among sexual minority men. *Cyberpsychol. Behav. Soc. Network.* 21, 149–156. doi: 10.1089/cyber.2017.0375

Hargreaves, D., and Tiggemann, M. (2003). The effect of "thin ideal" television commercials on body dissatisfaction and schema activation during

early adolescence. *J. Youth Adolesc.* 32, 367–373. doi: 10.1023/A:1024974015581

He, J., Chen, X., Fan, X., Cai, Z., and Huang, F. (2019). Is there a relationship between body mass index and academic achievement? A meta-analysis. *Public Health* 167, 111–124. doi: 10.1016/j.puhe.2018.11.002

He, J., Sun, S., Lin, Z., and Fan, X. (2020). Body dissatisfaction and sexual orientations: a quantitative synthesis of 30 years research findings. *Clin. Psychol. Rev.* 81:101896. doi: 10.1016/j.cpr.2020.101896

Hendrickse, J., Arpan, L. M., Clayton, R. B., and Ridgway, J. L. (2017). Instagram and college women's body image: Investigating the roles of appearance-related comparisons and intrasexual competition. *Comput. Hum. Behav.* 74, 92–100. doi: 10.1016/j.chb.2017.04.027

Higgins, J. P., Thompson, S. G., Deeks, J. J., and Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ* 327, 557–560. doi: 10.1136/bmj.327.7414.557

Hillege, S., Beale, B., and McMaster, R. (2006). Impact of eating disorders on family life: Individual parents' stories. *J. Clin. Nurs.* 15, 1016–1022. doi: 10.1111/j.1365-2702.2006.01367.x

Ho, S. S., Lee, E. W., and Liao, Y. (2016). Social network sites, friends, and celebrities: the roles of social comparison and celebrity involvement in adolescents' body image dissatisfaction. *Soc. Media+ Soc.* 2:2056305116664216. doi: 10.1177/2056305116664216

Hofstede, G. (1984). *Culture's Consequences: International Differences in Work-Related Values*. Thousand Oaks, CA: Sage.

Holland, G., and Tiggemann, M. (2016). A systematic review of the impact of the use of social networking sites on body image and disordered eating outcomes. *Body Image* 17, 100–110. doi: 10.1016/j.bodyim.2016.02.008

Holland, G., and Tiggemann, M. (2017). "Strong beats skinny every time": disordered eating and compulsive exercise in women who post fitspiration on Instagram. *Int. J. Eat. Disord.* 50, 76–79. doi: 10.1002/eat.22559

Howard, L. M., Heron, K. E., MacIntyre, R. I., Myers, T. A., and Everhart, R. S. (2017). Is use of social networking sites associated with young women's body dissatisfaction and disordered eating? A look at Black-White racial differences. *Body Image*, 23, 109–113. doi: 10.1016/j.bodyim.2017.08.008

Hutton, B., Salanti, G., Caldwell, D. M., Chaimani, A., Schmid, C. H., Cameron, C., et al. (2015). The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. *Ann. Intern. Med.* 162, 777–784. doi: 10.7326/M14-2385

Internet Live Stats. (2021). Retrieved from: https://www.internetlivestats.com/ (accessed January 13, 2021).

Jankowiak, W., Gray, P. B., and Hattman, K. (2008). Globalizing evolution: female choice, nationality, and perception of sexual beauty in China. *Cross-Cultural Res.* 42, 248–269. doi: 10.1177/1069397108317672

Jenkins, P. E., Hoste, R., Conley, C., Meyer, C., and Blissett, J. (2011b). Is being underweight associated with impairments in quality of life in the absence of significant eating disorder pathology. *Eat. Weight Disord.* 16, e61–e64. doi: 10.1007/BF03327523

Jenkins, P. E., Hoste, R. R., Meyer, C., and Blissett, J. M. (2011a). Eating disorders and quality of life: a review of the literature. *Clin. Psychol. Rev.* 31, 113–121. doi: 10.1016/j.cpr.2010.08.003

Johnson, J., Spitzer, R., and Williams, J. B. (2001). Health problems, impairment and illnesses associated with bulimia nervosa and binge eating disorder among primary care and obstetric gynaecology patients. *Psychol. Med.* 31, 1455–1466. doi: 10.1017/S0033291701004640

Joinson, A. (1999). Social desirability, anonymity, and Internet-based questionnaires. *Behav. Res. Methods Instrum. Comput.* 31, 433–438. doi: 10.3758/BF03200723

Kaplan, A. M., and Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons* 53, 59–68. doi: 10.1016/j.bushor.2009.09.003

Kärkkäinen, U., Mustelin, L., Raevuori, A., Kaprio, J., and Keski-Rahkonen, A. (2018). Do disordered eating behaviours have long-term health-related consequences? *Europ. Eat. Disord. Rev.* 26, 22–28. doi: 10.1002/erv.2568

Kaur, P., Dhir, A., Chen, S., and Rajala, R. (2016). Understanding online regret experience using the theoretical lens of flow experience. *Comput. Hum. Behav.* 57, 230–239. doi: 10.1016/j.chb.2015.12.041

Keel, P. K., and Forney, K. J. (2013). Psychosocial risk factors for eating disorders. *Int. J. Eat. Disord.* 46, 433–439. doi: 10.1002/eat.22094

Keery, H., Van den Berg, P., and Thompson, J. K. (2004). An evaluation of the Tripartite Influence Model of body dissatisfaction and eating disturbance with adolescent girls. *Body Image* 1, 237–251. doi: 10.1016/j.bodyim.2004.03.001

Kim, J. W., and Chock, T. M. (2015). Body image 2.0: associations between social grooming on Facebook and body image concerns. *Comput. Hum. Behav.* 48, 331–339. doi: 10.1016/j.chb.2015.01.009

Klaczynski, P. A., and Felmban, W. S. (2019). Thin idealization and causal attributions mediate the association between culture and obesity stereotypes: an examination of Chinese and American adolescents. *Br. J. Dev. Psychol.* 37, 14–32. doi: 10.1111/bjdp.12250

Koiranen, I., Keipi, T., Koivula, A., and Räsänen, P. (2019). Changing patterns of social media use? A population-level study of Finland. *Univ. Access Inform. Soc.* 19, 603–617. doi: 10.1007/s10209-019-00654-1

Latzer, Y., Spivak-Lavi, Z., and Katz, R. (2015). Disordered eating and media exposure among adolescent girls: the role of parental involvement and sense of empowerment. *Int. J. Adolesc. Youth* 20, 375–391. doi: 10.1080/02673843.2015.1014925

Leung, F., Wang, J., and Tang, C. W. (2004). Psychometric properties and normative data of the Eating Disorder Inventory among 12 to 18 year old Chinese girls in Hong Kong. *J. Psychosom. Res.* 57, 59–66. doi: 10.1016/S0022-3999(03)00506-3

Lin, L., Chu, H., and Hodges, J. S. (2017). Alternative measures of between-study heterogeneity in meta-analysis: reducing the impact of outlying studies. *Biometrics* 73, 156–166. doi: 10.1111/biom.12543

Mabe, A. G., Forney, K. J., and Keel, P. K. (2014). Do you "like" my photo? Facebook use maintains eating disorder risk. *Int. J. Eating Disord.* 47, 516–523. doi: 10.1002/eat.22254

Madge, C., Meek, J., Wellens, J., and Hooley, T. (2009). Facebook, social integration and informal learning at university:'It is more for socialising and talking to friends about work than for actually doing work'. *Learn. Media Technol.* 34, 141–155. doi: 10.1080/17439880902923606

Marengo, D., Longobardi, C., Fabris, M., and Settanni, M. (2018). Highly-visual social media and internalizing symptoms in adolescence: the mediating role of body image concerns. *Comput. Hum. Behav.* 82, 63–69. doi: 10.1016/j.chb.2018.01.003

Marsh, H. W., Hau, K.-T., Sung, R. Y., and Yu, C.-W. (2007). Childhood obesity, gender, actual-ideal body image discrepancies, and physical self-concept in Hong Kong children: cultural differences in the value of moderation. *Dev. Psychol.* 43, 647–662. doi: 10.1037/0012-1649.43.3.647

Moessner, M., Fassnacht, D. B., and Bauer, S. (2015). Online assessment of eating disorders: the clinical and research inventory for eating disorders (CR-EAT). *Mental Health Prevent.* 3, 170–177. doi: 10.1016/j.mhp.2015.08.001

Neumark-Sztainer, D., Wall, M., Story, M., and Standish, A. R. (2012). Dieting and unhealthy weight control behaviors during adolescence: associations with 10-year changes in body mass index. *J. Adolesc. Health* 50, 80–86. doi: 10.1016/j.jadohealth.2011.05.010

Niu, G., Sun, L., Liu, Q., Chai, H., Sun, X., and Zhou, Z. (2019). Selfie-posting and young adult women's restrained eating: the role of commentary on appearance and self-objectification. *Sex Roles* 82, 232–240. doi: 10.1007/s11199-019-01045-9

Obst, P., and Stafurik, J. (2010). Online we are all able bodied: online psychological sense of community and social support found through membership of disability-specific websites promotes well-being for people living with a physical disability. *J. Commun. Appl. Soc. Psychol.* 20, 525–531. doi: 10.1002/casp.1067

Perloff, R. M. (2014). Social media effects on young women's body image concerns: theoretical perspectives and an agenda for research. *Sex Roles* 71, 363–377. doi: 10.1007/s11199-014-0384-6

Pollack, L. O. (2017). *Examination of non-facial emotion recognition in adolescents with anorexia nervosa* (Doctoral dissertation), University of Missouri, Kansas City, MI. Retrieved from: https://hdl.handle.net/10355/61497

Przybylski, A. K., Murayama, K., DeHaan, C. R., and Gladwell, V. (2013). Motivational, emotional, and behavioral correlates of fear of missing out. *Comput. Hum. Behav.* 29, 1841–1848. doi: 10.1016/j.chb.2013.02.014

R Core Team (2020). *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing. Retrieved from: https://www.R-project.org/ (accessed May 30, 2020).

Rammstedt, B., Riemann, R., Angleitner, A., and Borkenau, P. (2004). Resilients, overcontrollers, and undercontrollers: the replicability of the three personality prototypes across informants. *Europ. J. Personality* 18, 1–14. doi: 10.1002/per.495

Ramsey, L. R., and Horan, A. L. (2018). Picture this: Women's self-sexualization in photos on social media. *Personal. Individ. Diff.* 133, 85–90. doi: 10.1016/j.paid.2017.06.022

Rodgers, R. F. (2016). The role of the "Healthy Weight" discourse in body image and eating concerns: an extension of sociocultural theory. *Eat. Behav.* 22, 194–198. doi: 10.1016/j.eatbeh.2016.06.004

Rodgers, R. F., DuBois, R., Frumkin, M. R., and Robinaugh, D. J. (2018). A network approach to eating disorder symptomatology: do desire for thinness and fear of gaining weight play unique roles in the network? *Body Image* 27, 1–9. doi: 10.1016/j.bodyim.2018.07.004

Rodgers, R. F., and Melioli, T. (2016). The relationship between body image concerns, eating disorders and internet use, part I: a review of empirical support. *Adoles. Res. Rev.* 1, 95–119. doi: 10.1007/s40894-015-0016-6

Rodgers, R. F., Slater, A., Gordon, C. S., McLean, S. A., Jarman, H. K., and Paxton, S. J. (2020). A biopsychosocial model of social media use and body image concerns, disordered eating, and muscle-building behaviors among adolescent girls and boys. *J. Youth Adolesc.* 49, 399–409. doi: 10.1007/s10964-019-01190-0

Ryding, F. C., and Kuss, D. J. (2019). The use of social networking sites, body image dissatisfaction, and body dysmorphic disorder: a systematic review of psychological research. *Psychol. Popular Media.* 9, 412–435. doi: 10.1037/ppm0000264

Santarossa, S. (2015). *#Socialmedia: Exploring the Associations of Social Networking Sites and Body Image, Self-Esteem, Disordered Eating and/or Eating Disorders and the Impact of a Media Literacy Intervention. (M.H.K.).* Ann Arbor, MI: University of Windsor (Canada). Retrieved from: https://search.proquest.com/docview/1730396652?accountid=12206 (accessed May 30, 2020). ProQuest Dissertations & Theses Global database. (1599625)

Santarossa, S., and Woodruff, S. J. J. S. M. S. (2017). # SocialMedia: Exploring the relationship of social networking sites on body image, self-esteem, and eating disorders. 3, 2056305117704407. doi: 10.1177/2056305117704407

Schaumberg, K., Welch, E., Breithaupt, L., Hübel, C., Baker, J. H., Munn-Chernoff, M. A., et al. (2017). The science behind the Academy for Eating Disorders' nine truths about eating disorders. *Eur. Eat. Disord. Rev.* 25, 432–450. doi: 10.1002/erv.2553

Schmidt, U., Adan, R., Böhm, I., Campbell, I. C., Dingemans, A., Ehrlich, S., et al. (2016). Eating disorders: the big issue. *Lancet Psychiatry* 3, 313–315. doi: 10.1016/S2215-0366(16)00081-X

Schreyer-Hoffman, G. (2020). *Understanding the Factors that Increase Disordered Eating and Eating Disorder Risk among Female Undergraduates.* ETD Collection for Fordham University. AAI27669065. Retrieved from: https://research.library.fordham.edu/dissertations/AAI27669065

Sedikides, C., Gaertner, L., and Toguchi, Y. (2003). Pancultural self-enhancement. *J. Personal. Soc. Psychol.* 84, 60–79. doi: 10.1037/0022-3514.84.1.60

Sharma, S. and Verma, H. V. (2018). "Social media marketing: evolution and change," in *Social Media Marketing*, eds S. G. Heggde G. içinde (Singapore: Palgrave Macmillan), 19–36. doi: 10.1007/978-981-10-5323-8_2

Slater, A., and Tiggemann, M. (2014). Media exposure, extracurricular activities, and appearance-related comments as predictors of female adolescents' self-objectification. *Psychol. Women Q.* 39, 375–389. doi: 10.1177/0361684314554606

Smith, A. R., Hames, J. L., and Joiner Jr, T. E. (2013). Status update: Maladaptive Facebook usage predicts increases in body dissatisfaction and bulimic symptoms. *J. Affect. Disord.* 149, 235–240. doi: 10.1016/j.jad.2013.01.032

Social Networking Fact Sheets, 2019 Social Networking Fact Sheets (2019). *PewResearchCenter.* Retrieved from: http://www.pewinternet.org/fact-sheets/social-networking-fact-sheet/

Spraggins, A. (2009). *Problematic Use of Online Social Networking Sites for College Students: Prevalence, Predictors, and Association With Well-Being.* Gainesville, FL: University of Florida.

Stice, E. (1994). Review of the evidence for a sociocultural model of bulimia nervosa and an exploration of the mechanisms of action. *Clin. Psychol. Rev.* 14, 633–661. doi: 10.1016/0272-7358(94)90002-7

Suplee, A. F. (2018). *An Examination of Social Media and the Tripartite Influence Model of Body Image Disturbance (Order No. AAI10193781).* Available from PsycINFO (1945236422; 2017-23162-083). Retrieved from https://proxy.library.upenn.edu/login?url=https://proxy.library.upenn.edu:2072/docview/1945236422?accountid=14707

Teo, N. S. Y., and Collinson, S. L. (2019). Instagram and risk of rumination and eating disorders: an Asian perspective. *Psychol. Popular Media Culture* 8, 491–508. doi: 10.1037/ppm0000205

Thelen, M. H., Farmer, J., Wonderlich, S., and Smith, M. (1991). A revision of the Bulimia Test: The BULIT—R. *Psychol. Assess.* 3:119. doi: 10.1037/1040-3590.3.1.119

Tiggemann, M. (2012). "Sociocultural perspectives on body image," in *Encyclopedia of Body Image and Human Appearance*, ed T. F. Cash (Elsevier Academic Press), 758–765. doi: 10.1016/B978-0-12-384925-0.00120-6

Tiggemann, M., and Miller, J. (2010). The Internet and adolescent girls' weight satisfaction and drive for thinness. *Sex Roles* 63, 79–90. doi: 10.1007/s11199-010-9789-z

Tiggemann, M., Slater, A., Bury, B., Hawkins, K., and Firth, B. (2013). Disclaimer labels on fashion magazine advertisements: Effects on social comparison and body dissatisfaction. *Body Image* 10, 45–53. doi: 10.1016/j.bodyim.2012.08.001

Tylka, T. L., and Sabik, N. J. (2010). Integrating social comparison theory and self-esteem within objectification theory to predict women's disordered eating. *Sex Roles* 63, 18–31. doi: 10.1007/s11199-010-9785-3

Valkenburg, P. M., and Peter, J. (2009). Social consequences of the Internet for adolescents: a decade of research. *Curr. Direct. Psychol. Sci.* 18, 1–5. doi: 10.1111/j.1467-8721.2009.01595.x

Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., and Sánchez-Meca, J. (2013). Three-level meta-analysis of dependent effect sizes. *Behav. Res. Methods* 45, 576–594. doi: 10.3758/s13428-012-0261-6

Vanden Abeele, M. M., Antheunis, M. L., Pollmann, M. M., Schouten, A. P., Liebrecht, C. C., Van Der Wijst, P. J., et al. (2018). Does Facebook use predict college students' social capital? A replication of Ellison, Steinfield, and Lampe's (2007) study using the original and more recent measures of Facebook use and social capital. *Commun. Stud.* 69, 272–282. doi: 10.1080/10510974.2018.1464937

Vannucci, A., and Ohannessian, C. M. (2019). Social media use subgroups differentially predict psychosocial well-being during early adolescence. *J. Youth Adoles.* 48, 1469–1493. doi: 10.1007/s10964-019-01060-9

Verduyn, P., Lee, D. S., Park, J., Shablack, H., Orvell, A., Bayer, J., et al. (2015). Passive Facebook usage undermines affective well-being: experimental and longitudinal evidence. *J. Exp. Psychol.* 144, 480. doi: 10.1037/xge0000057

Verduyn, P., Ybarra, O., Résibois, M., Jonides, J., and Kross, E. (2017). Do social network sites enhance or undermine subjective well-being? A critical review. *Soc. Issues Policy Rev.* 11, 274–302. doi: 10.1111/sipr.12033

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor Package. *J. Statist. Software* 36, 1–48. doi: 10.18637/jss.v036.i03

Viechtbauer, W., and Cheung, M. W. L. (2010). Outlier and influence diagnostics for meta-analysis. *Res. Synthesis Methods*, 1, 112–125. doi: 10.1002/jrsm.11

Walker, M., Thornton, L., De Choudhury, M., Teevan, J., Bulik, C. M., Levinson, C. A., et al. (2015). Facebook use and disordered eating in college-aged women. *J. Adolesc. Health* 57, 157–163. doi: 10.1016/j.jadohealth.2015.04.026

Wilksch, S. M., O'Shea, A., Ho, P., Byrne, S., and Wade, T. D. (2019). The relationship between social media use and disordered eating in young adolescents. *Int. J. Eat. Disord.* 53, 96–106. doi: 10.1002/eat.23198

Xue, K., and Yu, M. (2017). *New Media and Chinese Society (Vol. 5).* Singapore: Springer. doi: 10.1007/978-981-10-6710-5

Yao, L., Niu, G., and Sun, X. (2021). Body image comparisons on social networking sites and chinese female college students' restrained eating: the roles of body shame, body appreciation, and body mass index. *Sex Roles* 84, 465–476. doi: 10.1007/s11199-020-01179-1

Zeeni, N., Doumit, R., Abi Kharma, J., and Sanchez-Ruiz, M.-J. (2018). Media, technology use, and attitudes: associations with physical and mental well-being

in youth with implications for evidence-based practice. *Worldviews Evid. Based Nurs.* 15, 304–312. doi: 10.1111/wvn.12298

Zhang, Y., and Leung, L. (2015). A review of social networking service (SNS) research in communication journals from 2006 to 2011. *New Media Soc.* 17, 1007–1024. doi: 10.1177/1461444813520477

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The Reviewer XX declared a shared affiliation with several of the authors, JZ, YW, and QL, to the handling editor at time of review.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

APPENDIX A

Quality assessment.

References	Q1	Q2	Q3	Q4	Q5	Q6	Total	Quality
Santarossa (2015)	1	1	1	1	1	1	6	100%
Blassngame (2020a)	1	0	1	1	1	1	5	83.3%
Cohen et al. (2017)	0	0	1	1	0	1	4	66.7%
Rodgers et al. (2020)	1	0	1	1	1	1	5	83.3%
Suplee (2018)	1	0	1	1	1	1	5	83.3%
Ferguson et al. (2013)	0	0	1	1	1	1	4	66.7%
Latzer et al. (2015)	0	1	1	1	1	1	5	83.3%
Mabe et al. (2014)	0	0	1	1	1	1	4	66.7%
Acar et al. (2020)	0	0	1	1	1	1	4	66.7%
Pollack (2017)	1	0	1	1	1	1	4	66.7%
Walker et al. (2015)	1	1	1	1	1	1	6	100.0%
Griffiths et al. (2018a)	1	1	1	1	0	1	5	83.3%
Teo and Collinson (2019)	1	1	1	1	1	1	6	100.0%
Howard et al. (2017)	1	1	1	1	1	1	6	100.0%
Slater and Tiggemann (2014)	1	0	1	1	1	1	5	83.3%
Zeeni et al. (2018)	0	0	1	1	1	1	4	66.7%
Niu et al. (2019)	1	0	1	1	1	1	4	83.3%
Aparicio-Martinez et al. (2019)	0	1	1	1	1	1	5	83.3%
Griffiths et al. (2018b)	1	0	1	1	0	1	4	66.7%
Fardouly et al. (2020)	1	1	1	1	1	1	6	100.0%
Wilksch et al. (2019)	0	0	1	1	1	1	4	66.7%
Schreyer-Hoffman (2020)	1	0	1	1	1	1	4	83.3%

Q1 Sampling method: Was it representative of the population intended in the study?.
Q2 Was a response rate mentioned within the study? (Respond no if response rate was below 60%).
Q3 Was the measurement tool of disordered eating behaviors valid and reliable?
Q4 Was the data source primary or secondary?.
Q5 Was SNSs usage examined in the study in a valid and reliable way?
Q6 Was the relationship or association between disordered eating behaviors and SNSs explored?

12

Understanding Perceptions of Problematic Facebook Use

When People Experience Negative Life Impact and a Lack of Control

Justin Cheng
Facebook
jcheng@fb.com

Moira Burke
Facebook
mburke@fb.com

Elena Goetz Davis
Facebook
elenadavis@fb.com

ABSTRACT

While many people use social network sites to connect with friends and family, some feel that their use is problematic, seriously affecting their sleep, work, or life. Pairing a survey of 20,000 Facebook users measuring perceptions of problematic use with behavioral and demographic data, we examined Facebook activities associated with problematic use as well as the kinds of people most likely to experience it. People who feel their use is problematic are more likely to be younger, male, and going through a major life event such as a breakup. They spend more time on the platform, particularly at night, and spend proportionally more time looking at profiles and less time browsing their News Feeds. They also message their ... more frequently. While they are more likely to re-
... eactivate
... age their
... ent about
... eporting
... em, high-
... ology use
... atic Face-
... priate and
... ontrol.

Addictin

... orking sites;

ACM ISBN 978-1-4503-5970-2/19/05...$15.00
https://doi.org/10.1145/3290605.3300429

ACM Reference Format:
Justin Cheng, Moira Burke, and Elena Goetz Davis. 2019. Understanding Perceptions of Problematic Facebook Use: When People Experience Negative Life Impact and a Lack of Control. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland UK.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3290605.3300429

1 INTRODUCTION

Social network sites help people maintain social relationships [17, 31], drive civic engagement and collective action [35, 68], and support entrepreneurship [43]. But while many people derive benefit from online social networks, some feel that their use of such services is problematic. Studies of problematic use of the internet (e.g., [21, 102]) and social networks (e.g., [2, 61, 81]) note symptoms including preoccupation, loss of control, and negative impact on one's relationships, work performance, and life [40].

The present study focuses on perceived problematic Facebook use to understand its prevalence and its relation to different activities on the site, in order to inform design improvements that may reduce problematic use. We define "problematic Facebook use" as reporting a significant negative impact on sleep, relationships, or work or school performance and feeling a lack of control over site use, consistent with broad definitions from the academic literature [72, 81]. We do not use the term "addiction" because there is no agreed-upon criteria for diagnosis [8, 41, 89], and because diagnoses of clinical-level concerns would require more formal assessment (i.e., by a mental health professional) [55]. Instead, we focus on self-reported problematic use to understand differences across a broad population of users.

We pair a survey of 20,000 Facebook users in the U.S. measuring perceived problematic Facebook use with server logs of aggregated behavioral data for the previous four weeks, such as the amount of time respondents spent on the site and counts of interactions with close friends. In contrast to prior work that has relied on small-sample, survey-only analyses of problematic social network use [81], mostly among adolescents and young adults (e.g., [2, 30]), we use a larger, more diverse sample to study how perceptions of problematic use relate to actual on-site activity. By drawing data from both

Understanding Perceptions of Problematic Facebook Use

When People Experience Negative Life Impact and a Lack of Control

Justin Cheng
Facebook
jcheng@fb.com

Moira Burke
Facebook
mburke@fb.com

Elena Goetz Davis
Facebook
elenadavis@fb.com

ABSTRACT

While many people use social network sites to connect with friends and family, some feel that their use is problematic, seriously affecting their sleep, work, or life. Pairing a survey of 20,000 Facebook users measuring perceptions of problematic use with behavioral and demographic data, we examined Facebook activities associated with problematic use as well as the kinds of people most likely to experience it. People who feel their use is problematic are more likely to be younger, male, and going through a major life event such as a breakup. They spend more time on the platform, particularly at night, and spend proportionally more time looking at profiles and less time browsing their News Feeds. They also message their friends more frequently. While they are more likely to respond to notifications, they are also more likely to deactivate their accounts, perhaps in an effort to better manage their time. Further, they are more likely to have seen content about social media or phone addiction. Notably, people reporting problematic use rate the site as more valuable to them, highlighting the complex relationship between technology use and well-being. A better understanding of problematic Facebook use can inform the design of context-appropriate and supportive tools to help people become more in control.

CCS CONCEPTS

• **Human-centered computing → Social networking sites;**

KEYWORDS

Problematic use, Facebook

ACM ISBN 978-1-4503-5970-2/19/05...$15.00
https://doi.org/10.1145/3290605.3300429

ACM Reference Format:
Justin Cheng, Moira Burke, and Elena Goetz Davis. 2019. Understanding Perceptions of Problematic Facebook Use: When People Experience Negative Life Impact and a Lack of Control. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland UK.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3290605.3300429

1 INTRODUCTION

Social network sites help people maintain social relationships [17, 31], drive civic engagement and collective action [35, 68], and support entrepreneurship [43]. But while many people derive benefit from online social networks, some feel that their use of such services is problematic. Studies of problematic use of the internet (e.g., [21, 102]) and social networks (e.g., [2, 61, 81]) note symptoms including preoccupation, loss of control, and negative impact on one's relationships, work performance, and life [40].

The present study focuses on perceived problematic Facebook use to understand its prevalence and its relation to different activities on the site, in order to inform design improvements that may reduce problematic use. We define "problematic Facebook use" as reporting a significant negative impact on sleep, relationships, or work or school performance and feeling a lack of control over site use, consistent with broad definitions from the academic literature [72, 81]. We do not use the term "addiction" because there is no agreed-upon criteria for diagnosis [8, 41, 89], and because diagnoses of clinical-level concerns would require more formal assessment (i.e., by a mental health professional) [55]. Instead, we focus on self-reported problematic use to understand differences across a broad population of users.

We pair a survey of 20,000 Facebook users in the U.S. measuring perceived problematic Facebook use with server logs of aggregated behavioral data for the previous four weeks, such as the amount of time respondents spent on the site and counts of interactions with close friends. In contrast to prior work that has relied on small-sample, survey-only analyses of problematic social network use [81], mostly among adolescents and young adults (e.g., [2, 30]), we use a larger, more diverse sample to study how perceptions of problematic use relate to actual on-site activity. By drawing data from both

surveys and server logs, we reduce common-method bias, in which problematic outcomes and self-reported time online appear more associated than they are in reality [75].

Under this broad definition of problematic Facebook use – negative life impact and difficulty with control – we estimate (as an upper bound) that 3.1% of Facebook users in the US experience problematic use. They are more likely to be younger, male, and going through a major life event such as a breakup. After controlling for demographics, we find that people experiencing problematic use spend more time on the platform, particularly at night, and respond to a greater fraction of notifications. Contrary to stereotypes of people scrolling through endless content, people who experience problematic use spend proportionally less time in their News Feeds and more time browsing profiles, and message others more frequently. People reporting problematic use are also 2.6 times as likely to deactivate their accounts, perhaps as a way to control the time they spend on the site. They are also more likely to have viewed posts and comments about social media or phone addiction. And despite feeling that their use of the site has a negative impact in their lives, they rate Facebook as more valuable to them than do people in the non-problematic use group.

2 BACKGROUND

First, we review literature on problematic internet use and problematic Facebook use, and identify open questions about how individual differences and behaviors relate to the latter.

Problematic Internet and Facebook Use

Problematic internet use has been described as a set of symptoms including excessive amounts of time spent on the internet, a preoccupation with online activities or inability to control one's use, and adverse impact on one's social interactions and work or school performance [21]. Though academic researchers have described problematic internet use empirically, no formal clinical definition exists in either the Diagnostic and Statistical Manual of Mental Disorders [4] or the International Classification of Diseases [71]. There is also disagreement on whether such behaviors comprise a defined disorder [41] and whether research is pathologizing common behaviors [8]. Moreover, previous surveys that attempt to measure problematic internet use (e.g., [101]) have adopted inconsistent assessment criteria, leading to widely differing prevalence estimates [59]. These estimates have ranged from 1.5% to 8.2% in the US and Europe [98], to 0.3% to 38% internationally [22]. Nonetheless, there has been substantial academic and clinical interest in researching problematic internet use and related issues such as problematic Facebook use, problematic online gaming, and nomophobia (a fear of being out of mobile phone contact) [13, 61].

While there is debate on how problematic Facebook use should be measured [2, 41, 64] (or if it should be classified as an addiction [8]), a majority of survey instruments (e.g., [2]) include questions about lack of control, or a failure to abstain from the activity, and negative life impact, such as relationship conflict or reduced work or school performance [40, 60]. Other proposed symptoms from the behavioral addiction literature include salience, or how much one thinks about or engages in site use; tolerance, or needing increasing amounts of activity over time to achieve a desired effect; and mood modification and withdrawal, defined as a reliance on site use to reduce unpleasant feelings [40, 81]. Still, researchers have argued against using these symptoms as diagnostic criteria because of an absence of clinical studies [51]. Measures of symptoms such as tolerance that are adapted from diagnostic criteria for substance abuse may also not be appropriate when applied to technology use [8]. Consistent with survey instruments used in prior literature, this work focuses on problematic use as self-reporting both significant negative life impact and difficulty controlling Facebook use.

Prior literature suggests that symptoms of problematic internet use may be due to co-occuring problems [74] – individuals with problematic internet use tend to have other psychiatric disorders [57]. Past research has associated problematic internet or Facebook use with depression [52], lower happiness [14], worse academic performance [53], greater loneliness [82], and reduced relationship and life satisfaction [11, 32], though null results have also been reported [9]. Problematic internet behaviors may also arise from other individual differences. Previous work suggests that a preference for online social interaction may contribute to using the internet in problematic ways, especially when a person feels lonely or depressed [19], has low self-esteem [1, 50], or is neurotic or narcissistic [2, 66]. A fear of missing out ("FOMO"), more formally defined as "a pervasive apprehension that others might be having rewarding experiences from which one is absent" [76], might also contribute to problematic smartphone, internet, and social media use [15, 61, 69]. Problematic internet use has also been associated with structural differences in the brain [44].

Demographic Differences Related to Problematic Use

Gender. Evidence is mixed regarding whether men or women are more likely to experience problematic internet use. Previous work found that women tend to use Facebook more than men [34], and some studies indicate that women are more likely to experience problematic use [2, 5, 26] or report communication disturbance and phone obsession [10]. However, other studies showed a higher prevalence of problematic internet use among men [18, 30, 100]. Other work found no significant relation between gender and problematic internet use [11, 80, 88].

Age. Past research suggests that younger people may be more likely to experience problematic use because regions of the brain responsible for self-regulation are still developing in adolescence [86] and because they are more susceptible to negative peer influence [87]. Other work also found that internet addiction negatively correlates with age [33]. Correspondingly, a majority of previous studies of problematic use focus on these younger subpopulations (e.g., adolescents [5, 54] or college students [33, 56]). In the present work, we survey a wide range of Facebook users in the U.S. to better understand the relationship of both gender and age across a larger sample of people.

Because the existing literature is mixed on the relationship between gender and problematic use, and because little research has been done on problematic use across a wide range of ages, we pose the following research question:

RQ1: How does problematic use differ by gender and age?

Behaviors Associated With Problematic Use

Previous literature has also examined how specific behaviors relate to perceptions of problematic Facebook use. We discuss three main themes across behaviors: (1) excessive time spent, (2) connections and tie strength, and (3) loss of control. We also briefly examine the role of social narratives in shaping individual perceptions of problematic use.

Excessive time spent. Previous work has correlated time spent with both problematic internet [30, 33] and problematic Facebook use [46, 56]. Greater time spent has also been associated with social anxiety [83]. However, spending extended periods of time on the internet or on Facebook does not necessarily suggest problematic use [20]. Whereas generalized problematic internet use may involve displacing social connections, past research suggests that greater Facebook use may support people's relationships, depending on how they use it [17, 27, 96]. Moreover, research has also found that people spend substantial time on online social networks to maintain their offline social networks [60], and people who use Facebook several times a day have more close ties than people who do not [42]. Some work found a quadratic relationship between well-being and time spent online, with moderate use associated with improved well-being [77, 91].

Previous research has also linked problematic Facebook use to late-night use. It has been associated with both later bedtimes and rising times [2] and with insomnia [56, 92].

Overall, the relationship between time spent on Facebook and problematic use is unclear. Thus, we ask:

RQ2: How does time spent relate to problematic use?

Connections and tie strength. Past work has associated problematic Facebook use with poorer well-being [14, 81], so indicators of well-being may correlate negatively with problematic use. In particular, a large body of research suggests that interacting with close friends can lead to improvements in well-being, more so than interacting with acquaintances [7, 17, 93, 99]. If a person spends much of their time on Facebook interacting with acquaintances rather than close friends, this could influence their evaluation of the quality of the time she spends on the site, and their overall determination of whether their use is problematic. While little work on problematic internet use has focused on its relation to tie strength in online interactions, prior work noted higher levels of upward social comparison among people with more acquaintances in their Facebook friend graph [23]. Offline, people are more likely to underestimate others' difficult moments and overestimate their successes [49], and the effect online could be stronger among acquaintances, who may be less likely to know of each other's negative emotions. Thus, one might expect that if a person's Facebook network were denser and consisted of a greater ratio of strong to weak ties, they might experience improvements in well-being that buffer any negative impact from Facebook use.

Problematic use may also be associated with differences in friending and messaging behavior, but research is mixed. On one hand, individuals with low self-esteem may engage in friending more actively to compensate for a perceived deficiency [63]. On the other hand, teens who used Facebook to make friends reported reduced loneliness [90]. Prior work has associated instant messaging use with positive outcomes such as increased intimacy and improved relationship quality through increased self-disclosure [47, 94] and with negative outcomes such as problematic use [95]. As such:

RQ3: How do interactions with close friends on Facebook relate to problematic use?

Loss of control. Survey measures of problematic use commonly include lack of control [21]. We focus on two categories of Facebook activities that affect control: notifications and deactivation. Notifications may prompt people to use Facebook at times when they wouldn't have otherwise, thus reducing feelings of control by interrupting other tasks or in-person social interactions. In prior work, interruptions slow task completion [25], inhibit performance on complex tasks [85], and make it difficult to return to a previously interrupted task [70]. Previous research also found that notifications can cause inattention and hyperactivity, which in turn decreases productivity and subjective well-being [58].

RQ4: How do notifications differ between people who experience problematic use and those who don't?

Deactivation, or temporarily disabling one's account, is another method of control. Past work suggests that people may deactivate to focus during periods of high stress (e.g.,

before an exam), when they feel they spend too much time on Facebook, or to prevent others from interacting with their content while they are not online [6, 12].

RQ5: How do deactivation patterns relate to problematic use?

Social narratives. Previous research has shown that what people read or hear about can influence their beliefs [28, 65]. Reading an op-ed can result in substantial, long-term shifts in a person's policy opinions [24]. Further, previous qualitative work found that social narratives about smartphone addiction and its negative consequences can lead to people perceiving their own smartphone use negatively [62].

RQ6: How does reading about social media or smartphone addiction relate to perceptions of problematic use?

3 METHOD

To measure the prevalence of problematic Facebook use and the behaviors associated with it, we surveyed Facebook users in May 2018 and combined survey responses with server logs of the participants' activity on Facebook in the four weeks prior to them taking the survey. To protect participants' privacy, all data were de-identified, aggregated, and analyzed on Facebook's servers; no identifiable data were viewed by researchers. An internal board reviewed the research prior to the start of the study.

Participants. Participants (N=20,505; 62% female; mean age 44.5) were recruited via an ad on Facebook targeted at a random sample of people in the U.S. Compared to active Facebook users, respondents were on average 3.6 years older, 15% more likely to be female, had 20% more friends, and had owned their Facebook accounts for 1 year longer (all comparisons $p < 0.001$). To account for these differences, as well as general differences in site use due to demographics, we control for age, gender, friend count, and account tenure in all behavioral analyses below.

Problematic use survey. The survey contained questions about control and negative life impact adapted from the Internet Addiction Test [102], the Generalized Problematic Internet Use Scale 2 [64], and the Bergen Facebook Addiction Scale [2], see Table 1. The survey also asked "How valuable do you find the time you spend on Facebook?", "How meaningful are your interactions with people on Facebook?", and whether the respondent experienced any major life events in the past two months: 'moved to a new city', 'relationship breakup or divorce', 'lost job', 'new job', 'pregnancy or new family member', 'death of close friend or family', or 'personal injury or illness' [45]. Participants opted in to taking the survey; the survey stated "Some of these questions may be sensitive and all are optional; you may prefer not to answer and that's okay."

Defining problematic use. Based on the literature reviewed above, we defined problematic use as reporting both of the following:

(1) **Negative life impact** attributed to Facebook:
- Facebook hurts their relationships "very much," or
- They "very often" or "always" get less sleep because of Facebook, or
- Facebook hurts their work or school performance "greatly," or
- Facebook has a "very negative" impact on their lives

(2) and **problems with control or preoccupation:**
- "Very little or no control" over the time they spend on Facebook, or
- "Very" or "Extremely" concerned about missing posts from not logging in frequently enough

We require both components in our definition because voluntary choices (e.g., staying up late to use Facebook) may not be problematic if people still feel in control of those choices. We intentionally define this construct broadly, in contrast with stricter definitions proposed in the literature that require symptoms across multiple domains of functioning [2]. Thus, our estimate is likely to be an upper bound on the prevalence of problematic Facebook use, and more likely reflects risk of problematic use.

Measures of Potential Excessive Use

Time spent. We include the total amount of time participants spent on the site over the past four weeks as well as the number of distinct sessions because checking frequently throughout the day may indicate habitual behavior. A session is defined as distinct if it starts at least 60 seconds after a previous session ended. Similarly, we divided each day into 24 "hour bins" and counted the number of distinct bins in which a person had at least one session. Because of the association between problematic use and lack of sleep, we include the fraction of sessions that occur late at night (12 - 4 AM). We also include one measure from the survey in this analysis: how valuable respondents feel their time on Facebook is.

Furthermore, feelings of problematic use may be associated with how people spend their time on Facebook, so we include the proportion of time in News Feed (where they read friends' content and provide feedback), in Messenger (where they can have private conversations), on profiles, in groups, watching videos, and on Pages (which represent small businesses, public figures, or entertainment). Time in each of these areas was divided by the total time spent.

Measures of Connection and Tie Strength

Interactions with close friends. We defined "close friends" as a respondent's top 10 Facebook friends in terms of the number of mutual friends, among people with at least 100 friends. A

Survey items measuring problematic use
Negative life impact
How often do you get less sleep than you want because you're using Facebook? Never / Rarely / Sometimes / Very often / Always
Overall, how much does your use of Facebook hurt your relationships with others? Very slightly or not at all / A little / Moderately / Very much
To what extent does Facebook help or hurt your work or school performance? Helps greatly / Helps somewhat / Neither helps nor hurts / Hurts somewhat / Hurts greatly
Overall, do you feel like Facebook has had a positive or negative impact in your life? Very negative impact / Somewhat negative impact / Neither positive nor negative impact / Somewhat positive impact / Very positive impact
Control or preoccupation
How much control do you feel you have over the amount of time you spend on Facebook? Very little or no control / A little control / Some control / A lot of control / Complete control
How concerned are you about missing important posts on Facebook if you don't log in frequently enough? Not at all concerned / A little concerned / Somewhat concerned / Very concerned / Extremely concerned

Table 1: The problematic Facebook use survey included questions on negative life impact and control.

similar measure that defined closeness based on communication frequency and photo co-tags produced qualitatively similar results. In this study we measured the fraction of News Feed posts that respondents viewed that were produced by close friends, the fraction of messages to and from close friends, the fraction of profiles they viewed that were close friends, and their ego network density (or local clustering coefficient).

To additionally understand non-friend interactions, we measured the number of new friend requests sent in the past four weeks and the fraction of requests that were accepted, the fraction of non-friend profile views, and the fraction of comments written on non-friend posts.

Messaging. We include the number of messages participants sent in the past four weeks. All data were counts; no message text was viewed by researchers. Because the overall number of messages sent could simply be a proxy for overall time spent, we also included a normalized version, messages sent in four weeks divided by time spent on the site over four weeks. We also examined whether people sent more messages than they received from their friends. Because a person could send many short messages or fewer longer ones, we also included average message length.

Feedback received and given. Feedback in the form of likes and comments is a common component of social media models of tie strength [16, 36]. As feedback has been associated with improvements in well-being [17], it may also be linked to decreased problematic use. Thus, we measure the number of likes and comments received in the past four weeks normalized by the number of posts the participant wrote. We also measure the number of likes and comments participants gave, normalized by the number of posts they viewed.

Measures Related to Control

Notifications. We looked at the total number of push notifications received over the past four weeks, the fraction of notifications that a person responded to, and the mean time to response (in cases where there was a response).

Deactivation. We include whether the person deactivated their Facebook account in the four weeks prior to the survey. To see the survey, participants had to be currently active on the site, so these deactivations were temporary.

Measuring The Role of Social Narratives

To test if reading about social media or smartphone addiction is associated with feelings of problematic use, we analyzed posts and comments that participants viewed in the past four weeks, computing the fraction of posts and comments that included words relating to addiction (e.g., "addicted", "compulsive") as well as words relating to either social media or smartphones (e.g., "Facebook", "phone"). All analyses were done on de-identified data in aggregate; no post or comment text was viewed by researchers.

Demographic Variables

We include demographic variables including age and gender identity as covariates in our analyses, which are likely to affect both an individual's Facebook use and their perceptions of problematic use. We also include their friend count as a proxy for overall site engagement, and their account

tenure in days, to control for demographic differences based on when a person joined Facebook.

Method of Analysis

To understand how different experiences on Facebook are associated with reports of problematic use, we divided survey respondents into two groups: those who experience problematic use based on the definition above, and those who do not. For interpretability, we report results primarily as the relative differences in the means between the two groups based on a matched sample on age, gender, friend count, and account tenure (e.g., "people in the problematic use group spent 21.6% more time on Facebook than people in the non-problematic use group, all else being equal"). We performed coarsened exact matching [48], followed by a linear regression on the matched sample to compute the average treatment effect. To account for multiple comparison, we report Holm-corrected p-values. Comparing groups using logistic regressions controlling for age, gender, friend count, and account tenure on the entire data set (not matched samples) produces qualitatively similar results.

This method is correlational, so we cannot determine the causal relationship between survey measures of perceived problematic Facebook use and activities on the site. However, much of the existing research in this space is also correlational. By identifying associations, we can outline potential design implications and areas where additional research is needed to identify the causal direction.

4 RESULTS

Here, we examine the types of people that report problematic use, and explore how activity on Facebook relates to problematic use with respect to (1) potential excessive use, (2) connections and tie strength, (3) a loss of control, and (4) social narratives about addiction.

Who experiences problematic use?

Based on our definition for problematic use – experiencing a negative life outcome attributed to Facebook as well as a lack of control – 3.1% of Facebook users in the U.S. experience problematic use. This estimate has been weighted by age, gender, and time spent to account for selection bias among survey participants. Because of a lack of consensus in prior literature about how to define problematic use, we include the two most common criteria – a negative life outcome and lack of control. This is less restrictive than some models (e.g., those that require multiple negative life outcomes, mood modification, or tolerance). Therefore, our estimate of 3.1% is an upper bound compared to other definitions with stricter criteria, but this broader definition allows us to make design recommendations more broadly.

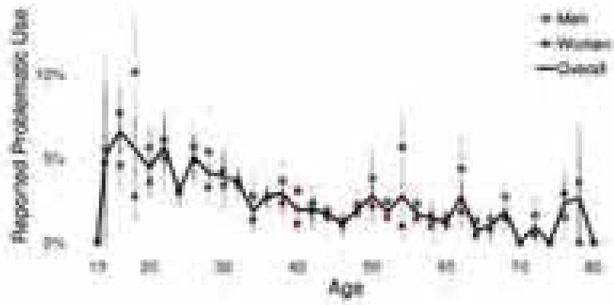


Figure 1: The prevalence of reported problematic use is highest among teens and young adults. Men are also more likely than women to report experiencing problematic use.

Age. Answers to Research Question 1, the prevalence of problematic use by age and gender, are presented in Figure 1. Perceptions of problematic use vary by age, with the prevalence highest among teens and young adults. People under the age of 25 were almost twice as likely as other age groups to experience problematic use (Cohen's $d = 0.13$, $p < 0.001$). This is consistent with previous research showing that younger people have more difficulty with self-regulation [86] and thus may be more prone to problematic use.

Gender. Men are 1.4x as likely as women to report experiencing problematic use ($d = 0.05$, $p < 0.001$, Figure 1). Still, the women in our sample spent 16% more time than the men on Facebook, suggesting that the relationship between time spent and problematic use is likely mediated by other factors, including motivations for use [79].

Major life events. Participants reported on the survey major life events that had happened in the past two months, and many of them were significantly associated with perceived problematic use (Figure 2). People who had recently gone through a breakup were 2.4x as likely to report that their use of Facebook was problematic. Similarly, a person who had recently moved to a new city was approximately 2x as likely to report problematic use.

Friend count and account tenure. Participants who reported experiencing problematic use had 29% more friends ($p < 0.001$), and had owned their Facebook accounts for about ten fewer months ($p < 0.001$). We control for these variables in all subsequent analyses.

Potential Excessive Use

Time spent. Consistent with prior literature, people who reported problematic use spent significantly more time on Facebook than people who did not (Figure 3a, Research Question 2). They spent 21.6% more time on the site ($d = 0.28$, $p < 0.001$), had 13.5% more distinct sessions ($d = 0.16$, $p < 0.05$),

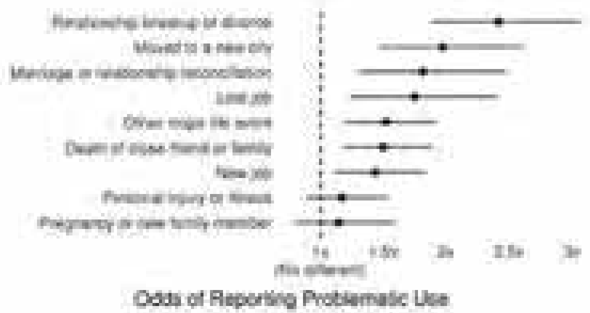


Figure 2: Having a major life event (e.g., a breakup) in the past two months is associated with an increased likelihood of reporting problematic use.

and had sessions in more distinct "hour bins" each day (d = 0.24, p < 0.001). They also spent a greater fraction of their sessions late at night (d = 0.16, p < 0.001), consistent with their increased likelihood of reporting sleep problems.

Despite reporting problems that they attributed to their Facebook use, individuals in the problematic use group found the time they spent on Facebook as 9.1% more valuable than people in the non-problematic use group (d = 0.24, p < 0.001). One interpretation is cognitive dissonance: a person justifies the extra time he or she spends on Facebook by thinking that it is more valuable. However, as we later show, there were no differences between groups in how meaningful they rated their interactions on the site. If cognitive dissonance explained the findings, we would expect people in the problematic use group to also rate their interactions as more meaningful. An alternative interpretation is that problematic use has both good and bad aspects to it – a person feels that they get value from Facebook, but may feel overly reliant on it or that they lack control.

The way they spent their time on the site also differed (Figure 3b). Compared to the non-problematic use group, people reporting problematic use spent a smaller proportion of their time viewing content on News Feed (-7.7%, d = 0.18, p < 0.001), and a greater proportion on profiles, both their own and others' (17.9%, d = 0.15, p < 0.01). They were no different in the proportion of their time they spent in Messenger, groups, videos, or Pages (n.s.).

Because of the greater proportion of time people experiencing problematic use spent on profiles, we conducted several post-hoc analyses to better understand if they were using profiles any differently. People in the problematic use group were no more likely to be viewing their own profile or a friend's (n.s.). Examining the fraction of profile views coming directly after a previous profile view, they were also no more likely to serially "hop" from profile to profile (n.s.). In addition, as profile pages include a link to message the

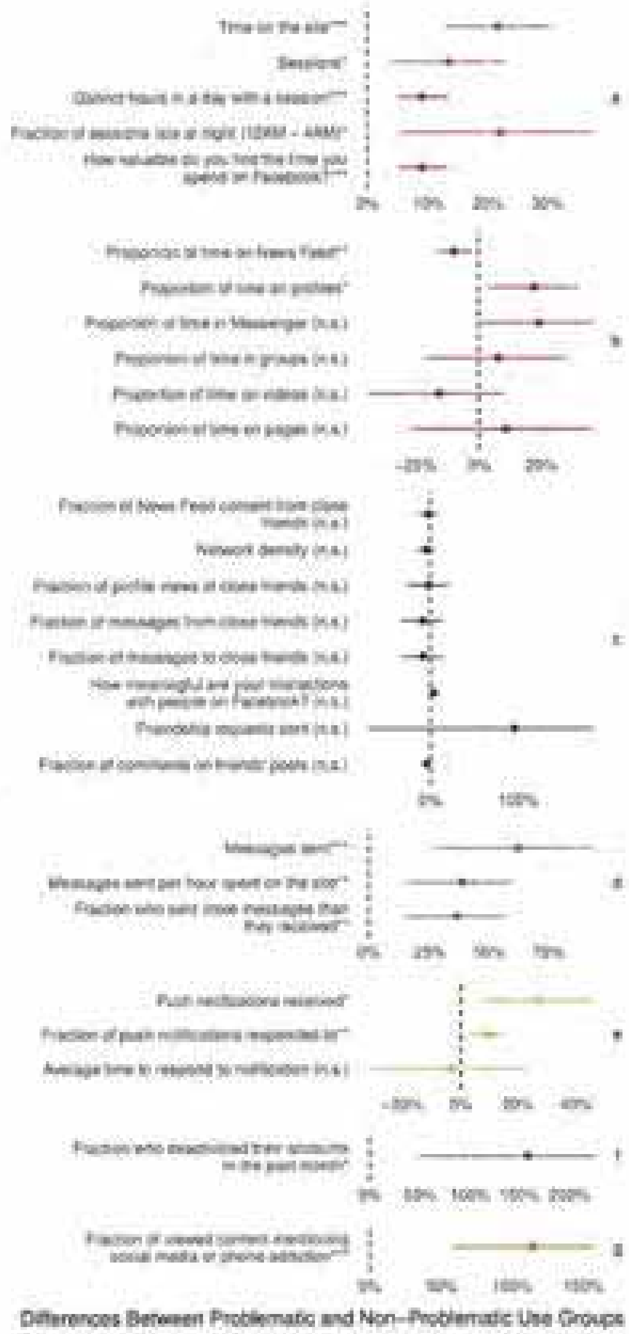


Figure 3: Relative differences between people who report experiencing problematic use and those who do not, matched on age, gender, friend count, and tenure. For example, people reporting problematic use spent 21.6% more time on Facebook than people who did not. Bars represent 99% bootstrapped confidence intervals. All p-values are Holm-corrected to account for multiple comparisons.
*** p < 0.05, ** p < 0.01, *** p < 0.001, n.s. not significant**

profile owner, the increased time spent on profiles may be partially due to people messaging others more (we discuss messaging in more detail later in this section). Indeed, a regression analysis revealed that the number of messages sent, number of friends messaged, and whether a person reported problematic use were all significant predictors of time spent viewing profiles ($p < 0.01$).

Connection and Tie Strength

Interactions with close friends. As strong-tie interactions have been associated with improved well-being, we expected that problematic use would be associated with fewer interactions with close friends and more interactions with weak ties. However, we found that people in the problematic use group were no different than people in the non-problematic use group in terms of the proportion of content they viewed from close friends, their network density, the proportion of close friends' profiles they viewed, and the proportion of messages they sent or received from close friends (Figure 3c, Research Question 3). They were also no different in how meaningful they said their interactions on the site were.

Problematic use was also not associated with people seeking out interactions outside of their friend networks: there were no group differences with respect to the frequency of sending friend requests, likelihood of friend requests being accepted, fraction of profile views that were of non-friends, or fraction of comments on non-friend posts (n.s.).

Synchronous messaging. While people experiencing problematic use do not spend proportionally more time messaging others, they still sent 62.7% more messages than those who are not experiencing problematic use ($d = 0.20$, $p < 0.001$) (Figure 3d), despite spending only 21.6% more time overall on Facebook. Normalizing by the amount of time spent on the site, they sent 38.7% more messages per hour ($d = 0.19$, $p < 0.001$). They were also 36.7% more likely to have sent more messages than they received ($d = 0.19$, $p < 0.01$). There were no differences in the mean number of words per message they sent or received, suggesting that these differences are not due to longer messages being split up into a series of smaller ones.

Overall, our findings on messaging activity and time spent contradict an image of people experiencing problematic use because of hours of unintentional scrolling or serially watching videos. Instead, they paint a picture of people spending more time browsing profiles and messaging others.

Feedback received and given. There were no significant differences between the problematic and non-problematic use groups in the number of likes per post or comments per post that people received, or in the number of likes or comments people gave per post that they viewed (not shown).

Control

Notifications. People reporting problematic use received 27.4% more notifications than people who did not report problematic use ($d = 0.15$, $p < 0.05$), and responded to a greater fraction of these notifications ($d = 0.18$, $p < 0.01$, Figure 3e, Research Question 4). In particular, they were more likely to respond to notifications when they were about replies to comments they had made ($d = 0.18$, $p < 0.05$). They did not respond to notifications any more quickly (n.s.).

The correlational data do not allow us to determine if notifications contribute to feelings of problematic use, or if the differences in notification volume and likelihood of responding reflect different levels of engagement and friend activity (e.g., more friends sharing content).

Deactivation. People in the problematic use group were 2.6x as likely to have deactivated their accounts in the past four weeks ($d = 0.16$, $p < 0.05$), compared to people in the non-problematic use group (Figure 3f, Research Question 5). These deactivations were temporary as respondents had to be using Facebook to be recruited for the survey, so the true number of deactivations among all individuals experiencing problematic use may have been larger. Previous research has described deactivation as a risk-reduction strategy [12] and way to focus [6], and that may be the case here: people who feel out of control about their Facebook use may deactivate their accounts to stop notifications, prevent themselves from habitually checking up on friends, or generally take a break from the site.

When people deactivate their accounts, Facebook requires that they provide a reason from a list of options, such as "I get too many emails, invitations, and requests from Facebook" or "I spend too much time using Facebook." There were no significant differences between groups in deactivation reasons, suggesting that people have similar reasons for deactivating, even if their use of the site isn't problematic. As we discuss below, designers of social network sites may want to offer more granular controls than deactivation to allow people to better manage their time, prevent interruptions, and break problematic habits.

Social Narratives

Participants experiencing problematic use were 2.1x as likely to have viewed posts and comments about social media or phone addiction ($d = 0.22$, $p < 0.001$, Figure 3g, Research Question 6). On one hand, people who experience problematic use may be more likely to look up content about addiction or know others who also experience problematic use. On the other hand, people who are exposed to discussion about social media or smartphone addiction may be more likely to think about these problems in their own lives.

5 DISCUSSION AND CONCLUSION

Summary. Approximately 3% of Facebook users in the U.S. report feeling like Facebook contributes to problems with their sleep, work, or relationships and that their Facebook use is difficult to control. Understanding their experiences on the platform can help designers develop supportive and context-appropriate tools to reduce negative impact associated with problematic use. This study presents several key differences between people reporting problematic use and those who do not, including greater time spent, particularly late at night; responding to a larger fraction of notifications; spending a greater proportion of time browsing friends' profiles; being more likely to deactivate; sending more messages than one receives; and reading more content about technology addiction. Demographic factors also play a role: men and younger people were more likely to feel that their use of Facebook was problematic, as were people who had gone through recent major life events such as breakups or moves.

Despite feeling like there were areas of their lives that were negatively impacted by Facebook use, people in the problematic use group also rated Facebook as more valuable in their lives than did people in the non-problematic use group, demonstrating that the technology is not uniformly beneficial or harmful. As designers, we should identify ways to help people avoid problematic use so that people can continue to get that value. People in the problematic group were nearly three times as likely to deactivate their accounts, which suggests they were attempting to gain more control over their time on the site, but deactivation cuts off access to that value. Later, we discuss design implications that may be more useful and flexible than deactivation.

Major life events. Major life events such as breakups or moves were associated with higher rates of problematic use, but the causal direction is unclear. Breakups could cause people to use technology in different ways, such as sending more messages to friends seeking support, or surveilling an ex's profile. They could also cause people to view their lives through a lens that makes other activities, such as technology use, seem problematic. Or, a major life event could be associated with a change in routine, and thus could be a vulnerable time for problematic patterns to be strengthened, when people have more time on their hands, are feeling upset or less social, or have something important they want to talk about through social media. Reverse causation is also possible: problematic technology use could lead to major life events; if technology use negatively affects sleep, relationships, or work performance, it could lead to a breakup, job loss, or move. Here, we do find an association between major life events and changes in behavior across both problematic and non-problematic use groups. In our data, major life events predicts more message-sending in both groups ($p < 0.001$). But while major life events do play a role in problematic use, they do not entirely account for the differences in behavior associated with problematic use. People experiencing problematic use still send more messages, even after controlling for whether they had a major life event in the past two months ($p < 0.001$).

Moderation in use. As we show in the present study, people who feel like they have a problem are more likely to deactivate their accounts. However, total avoidance may not be the best solution for everyone. People with problematic use are also more likely to report that their use of Facebook is more valuable, and our findings do not support the interpretation this this is due to cognitive dissonance or rationalization. Studies further show that moderate social media use results in more positive well-being outcomes than no social media use [77, 91]. Nonetheless, while moderate, controlled use may be the most appropriate recommendation for the general population, abstinence from problematic applications may still be warranted for individuals experiencing clinical-level concern with their internet behaviors [73, 103].

Design Implications

These findings suggest multiple opportunities for design, not just on Facebook, but communication platforms more generally. First, the data suggest the need to provide people with more granular options than deactivation to take a break from social media. Designers may want to promote alternative options to increase control and provide for uninterrupted time, such as turning off push notifications, especially at bedtime. Because there were no differences in the reasons for deactivation between the problematic and non-problematic use groups, design changes such as these could be relevant and beneficial for all users seeking a temporary break, not only those experiencing problematic use.

Several technology companies announced new features in 2018 to help people better manage interruptions [3, 38, 78]. Figure 4 shows Facebook's new time management tools, which were informed by this research. The tools include a dashboard to visualize time spent, a time-based reminder to take a break, and options to control or mute notifications.

Additional research is needed to understand what kinds of notifications people find most beneficial, so that designers can better prioritize and filter notifications. Teens and young adults were more likely to report experiencing problematic use, so designers may want to consider different control settings specifically for younger people.

Problematic use was higher among people experiencing certain major life events, including breakups and moves. These kinds of events are also associated with increases in depression [67]. Social media platforms could provide additional support for managing these life transitions.

Figure 6: Facebook's time management tools provide timers and reminders for people to take a break and options to mute notifications.

Limitations and Future Research

This method of pairing cross-sectional survey and behavioral data has several limitations. The analysis is correlational rather than causal; we can only report associations between perceived problematic use and site activities but do not know whether those activities cause feelings of problematic use, whether a person's propensity for problematic use causes those activity patterns, or whether something else like a major life event causes both perceptions of problematic use and site activity. We make design recommendations that we hypothesize will have positive outcomes, but further research is necessary to understand their impact on problematic use.

Though the present data come from Facebook, other smartphone apps and communication platforms may present similar opportunities to study problematic use. Notifications, browsing feeds of content, and channels for messaging are common across platforms. However, platforms differ in network composition, communication synchronicity, media type (e.g., images versus text posts), and motivations for use. For instance, some studies suggest that visual media provide more gratification than text media, and so may be perceived to be more compelling for increased use [29]. How these differences relate to problematic use remains future work.

More research is also necessary to understand how other factors may contribute to problematic use. For example, the popularity of social media and instant messaging has created pressure to always be available, particularly among youth [39]. While we found mixed evidence for this—people reporting problematic use responded to a larger fraction of notifications but did not respond any quicker—additional qualitative work could probe more deeply into the connection between availability expectations and problematic use. Upward social comparison may also lead to problematic use, especially among teens as peer influence is much stronger in adolescence than in adulthood [87].

Additional work is also necessary to better interpret the differences that we observed. For instance, though we found that people reporting problematic use spent proportionally more time viewing profiles, profile viewing is associated with both positive and negative outcomes. People who use profile pages more may be using Facebook primarily to keep up with friends they do not see as often, and this greater awareness of what others are doing can increase feelings of closeness [16]. However, spending time viewing profiles of acquaintances also lowers self-esteem [97] (though viewing one's own profile instead increases self-esteem [37]). Understanding the causal pathway between profile viewing and problematic use, if any exists, remains future work.

Several methodological limitations exist. For example, we only log time spent when the Facebook app is active on a person's screen, but do not know if they are looking at the screen the whole time; our measures of closeness may not necessarily identify every individual's closest friends. There are also selection biases among survey participants. People who have permanently quit Facebook are missing from the sample, so our statistics related to account deactivation are likely underestimates. We also do not know the relationship between problematic use and account deletion. Surveys outside of Facebook (e.g., [6]) are useful for understanding the motivations of people who have left the platform permanently. There may be other sources of response bias: people who stay up late may be more willing to complete surveys. The data are U.S.-centric and reports of problematic use and associated behaviors may differ internationally based on cultural differences, mobile broadband adoption, and norms. For example, time spent on social media varies by country. In 2017, people in the Philippines spent almost twice as much time on social media as people in the U.S. [84]. Additional international research is needed.

This research was quantitative. While it includes granular information about the kinds of activities that are associated with problematic use, we need additional qualitative research to better understand why people with problematic use engage with technology in the ways that they do and what would best help them gain control.

These challenges related to problematic use are not specific to Facebook—many of the findings in the present study generalize to other social media and smartphone technology more broadly. As researchers and designers we should continue to address the serious challenges that people face with technology in order to ensure it best serves its role in supporting people's lives.

ACKNOWLEDGMENTS

The authors would like to thank Bethany de Gant, Jennifer Guadagno, Alex Dow, Lada Adamic, and Robert Kraut for feedback and assistance with this work.

REFERENCES

[1] Cecilie Schou Andreassen, Ståle Pallesen, and Mark D Griffiths. 2017. The relationship between addictive use of social media, narcissism, and self-esteem: Findings from a large national survey. *Addictive Behaviors* (2017).

[2] Cecilie Schou Andreassen, Torbjørn Torsheim, Geir Scott Brunborg, and Ståle Pallesen. 2012. Development of a Facebook addiction scale. *Psychological Reports* (2012).

[3] Apple. 2018. iOS 12 introduces new features to reduce interruptions and manage Screen Time. https://nr.apple.com/dE4W3o6n3T.

[4] American Psychiatric Association. 2013. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®).*

[5] Fanni Bányai, Ágnes Zsila, Orsolya Király, Aniko Maraz, Zsuzsanna Elekes, Mark D Griffiths, Cecilie Schou Andreassen, and Zsolt Demetrovics. 2017. Problematic social media use: Results from a large-scale nationally representative adolescent sample. *PLoS One* (2017).

[6] Eric PS Baumer, Phil Adams, Vera D Khovanskaya, Tony C Liao, Madeline E Smith, Victoria Schwanda Sosik, and Kaiton Williams. 2013. Limiting, leaving, and (re) lapsing: an exploration of facebook non-use practices and experiences. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems.*

[7] Katherine Bessiere, Sara Kiesler, Robert Kraut, and Bonka S Boneva. 2008. Effects of Internet use and social resources on changes in depression. *Information, Community & Society* (2008).

[8] Joël Billieux, Adriano Schimmenti, Yasser Khazaal, Pierre Maurage, and Alexandre Heeren. 2015. Are we overpathologizing everyday life? A tenable blueprint for behavioral addiction research. *Journal of Behavioral Addictions* (2015).

[9] Roberta Biolcati, Giacomo Mancini, Virginia Pupi, and Valeria Mugheddu. 2018. Facebook Addiction: Onset Predictors. *Journal of Clinical Medicine* (2018).

[10] Agata Blachnio and Aneta Przepiorka. 2018. Be Aware! If You Start Using Facebook Problematically You Will Feel Lonely: Phubbing, Loneliness, Self-esteem, and Facebook Intrusion. A Cross-Sectional Study. *Social Science Computer Review* (2018).

[11] Agata Blachnio, Aneta Przepiorka, and Igor Pantic. 2016. Association between Facebook addiction, self-esteem and life satisfaction: A cross-sectional study. *Computers in Human Behavior* (2016).

[12] danah boyd. 2010. Risk reduction strategies on Facebook.

[13] Nicola Luigi Bragazzi and Giovanni Del Puente. 2014. A proposal for including nomophobia in the new DSM-V. *Psychology Research and Behavior Management* (2014).

[14] Julia Brailovskaia, Tobias Teismann, and Jürgen Margraf. 2018. Physical activity mediates the association between daily stress and Facebook Addiction Disorder (FAD)–A longitudinal approach among German students. *Computers in Human Behavior* (2018).

[15] Sarah L Buglass, Jens F Binder, Lucy R Betts, and Jean DM Underwood. 2017. Motivators of online vulnerability: The impact of social network site use and FOMO. *Computers in Human Behavior* (2017).

[16] Moira Burke and Robert E Kraut. 2014. Growing closer on facebook: changes in tie strength through social network site use. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems.*

[17] Moira Burke and Robert E Kraut. 2016. The relationship between Facebook use and well-being depends on communication type and tie strength. *Journal of Computer-Mediated Communication* (2016).

[18] Emre Çam and Onur Isbulan. 2012. A New Addiction for Teacher Candidates: Social Networks. *Turkish Online Journal of Educational Technology* (2012).

[19] Scott E Caplan. 2003. Preference for online social interaction: A theory of problematic Internet use and psychosocial well-being. *Communication Research* (2003).

[20] Scott E Caplan. 2005. A social skill account of problematic Internet use. *Journal of Communication* (2005).

[21] Hilarie Cash, Cosette D Rae, Ann H Steel, and Alexander Winkler. 2012. Internet addiction: A brief summary of research and practice. *Current Psychiatry Reviews* (2012).

[22] Kaustav Chakraborty, Debashis Basu, KG Kumar, et al. 2010. Internet addiction: consensus, controversies, and the way ahead. *East Asian Archives of Psychiatry* (2010).

[23] Hui-Tzu Grace Chou and Nicholas Edge. 2012. "They are happier and having better lives than I am": the impact of using Facebook on perceptions of others' lives. *Cyberpsychology, Behavior, and Social Networking* (2012).

[24] Alexander Coppock, Emily Ekins, David Kirby, et al. 2018. The Long-lasting Effects of Newspaper Op-Eds on Public Opinion. *Quarterly Journal of Political Science* (2018).

[25] Mary Czerwinski, Edward Cutrell, and Eric Horvitz. 2000. Instant messaging and interruption: Influence of task type on performance. In *Proceedings of the Australian Computer-Human Interaction Conference.*

[26] Rozane De Cock, Jolien Vangeel, Annabelle Klein, Pascal Minotte, Omar Rosas, and Gert-Jan Meerkerk. 2014. Compulsive use of social networking sites in Belgium: Prevalence, profile, and the role of attitude toward work and school. *Cyberpsychology, Behavior, and Social Networking* (2014).

[27] Fenne Große Deters and Matthias R Mehl. 2013. Does posting Facebook status updates increase or decrease loneliness? An online social networking experiment. *Social Psychological and Personality Science* (2013).

[28] Morton Deutsch and Harold B Gerard. 1955. A study of normative and informational social influences upon individual judgment. *Journal of Abnormal and Social Psychology* (1955).

[29] Áine Dunne, Margaret-Anne Lawlor, and Jennifer Rowley. 2010. Young people's use of online social networking sites–a uses and gratifications perspective. *Journal of Research in Interactive Marketing* (2010).

[30] Tony Durkee, Michael Kaess, Vladimir Carli, Peter Parzer, Camilla Wasserman, Birgitta Floderus, Alan Apter, Judit Balazs, Shira Barzilay, Julio Bobes, et al. 2012. Prevalence of pathological internet use among adolescents in Europe: demographic and social factors. *Addiction* (2012).

[31] Nicole B Ellison, Charles Steinfield, and Cliff Lampe. 2007. The benefits of Facebook "friends": Social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication* (2007).

[32] Rachel A Elphinston and Patricia Noller. 2011. Time to face it! Facebook intrusion and the implications for romantic jealousy and relationship satisfaction. *Cyberpsychology, Behavior, and Social Networking* (2011).

[33] Tania Fernández-Villa, Antonio J Molina, Miguel García-Martín, Javier Llorca, Miguel Delgado-Rodríguez, and Vicente Martín. 2015. Validation and psychometric analysis of the Internet Addiction Test in Spanish among college students. *BMC Public Health* (2015).

[34] Sarah Katherine Foregger. 2008. *Uses and gratifications of Facebook.com.*

[35] Homero Gil de Zúñiga, Nakwon Jung, and Sebastián Valenzuela. 2012. Social media use for news and individuals' social capital, civic engagement and political participation. *Journal of Computer-Mediated Communication* (2012).

[36] Eric Gilbert and Karrie Karahalios. 2009. Predicting tie strength with social media. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems.*

[37] Amy L Gonzales and Jeffrey T Hancock. 2011. Mirror, mirror on my Facebook wall: Effects of exposure to Facebook on self-esteem.

Cyberpsychology, Behavior, and Social Networking (2011).
[38] Google. 2018. Digital Wellbeing. https://wellbeing.google/.
[39] David Greenfield. 2011. The addictive properties of internet usage. *Internet addiction: A handbook and guide to evaluation and treatment* (2011).
[40] Mark Griffiths. 2005. A 'components' model of addiction within a biopsychosocial framework. *Journal of Substance Use* (2005).
[41] Mark D Griffiths. 2012. Facebook addiction: concerns, criticism, and recommendations—a response to Andreassen and colleagues. *Psychological Reports* (2012).
[42] Keith Hampton, Lauren Sessions Goulet, Lee Rainie, and Kristen Purcell. 2011. Social networking sites and our lives. *Pew Internet & American Life Project* (2011).
[43] Lisa Harris and Alan Rae. 2009. Social networks: the future of marketing for small business. *Journal of Business Strategy* (2009).
[44] Qinghua He, Ofir Turel, and Antoine Bechara. 2017. Brain anatomy alterations associated with Social Networking Site (SNS) addiction. *Scientific Reports* (2017).
[45] Thomas H Holmes and Richard H Rahe. 1967. The social readjustment rating scale. *Journal of Psychosomatic Research* (1967).
[46] Fu-Yuan Hong, Der-Hsiang Huang, Hung-Yu Lin, and Su-Lin Chiu. 2014. Analysis of the psychological traits, Facebook usage, and Facebook addiction model of Taiwanese university students. *Telematics and Informatics* (2014).
[47] Yifeng Hu, Jacqueline Fowler Wood, Vivian Smith, and Nalova Westbrook. 2004. Friendships through IM: Examining the relationship between instant messaging and intimacy. *Journal of Computer-Mediated Communication* (2004).
[48] Stefano M Iacus, Gary King, and Giuseppe Porro. 2012. Causal inference without balance checking: Coarsened exact matching. *Political Analysis* (2012).
[49] Alexander H Jordan, Benoît Monin, Carol S Dweck, Benjamin J Lovett, Oliver P John, and James J Gross. 2011. Misery has more company than people think: Underestimating the prevalence of others' negative emotions. *Personality and Social Psychology Bulletin* (2011).
[50] Yaniv Kanat-Maymon, Lian Almog, Rinat Cohen, and Yair Amichai-Hamburger. 2018. Contingent self-worth and Facebook addiction. *Computers in Human Behavior* (2018).
[51] Daniel Kardefelt-Winther, Alexandre Heeren, Adriano Schimmenti, Antonius van Rooij, Pierre Maurage, Michelle Carras, Johan Edman, Alexander Blaszczynski, Yasser Khazaal, and Joël Billieux. 2017. How can we conceptualize behavioural addiction without pathologizing common behaviours? *Addiction* (2017).
[52] Kyunghee Kim, Eunjung Ryu, Mi-Young Chon, Eun-Ja Yeun, So-Young Choi, Jeong-Seok Seo, and Bum-Woo Nam. 2006. Internet addiction in Korean adolescents and its relation to depression and suicidal ideation: a questionnaire survey. *International Journal of Nursing Studies* (2006).
[53] Paul A Kirschner and Aryn C Karpinski. 2010. Facebook® and academic performance. *Computers in Human Behavior* (2010).
[54] Chih-Hung Ko, Ju-Yu Yen, Cheng-Chung Chen, Sue-Huei Chen, and Cheng-Fang Yen. 2005. Gender differences and related factors affecting online gaming addiction among Taiwanese adolescents. *The Journal of Nervous and Mental Disease* (2005).
[55] Chih-Hung Ko, Ju-Yu Yen, Sue-Huei Chen, Ming-Jen Yang, Huang-Chi Lin, and Cheng-Fang Yen. 2009. Proposed diagnostic criteria and the screening and diagnosing tool of Internet addiction in college students. *Comprehensive Psychiatry* (2009).
[56] Mustafa Koc and Seval Gulyagci. 2013. Facebook addiction among Turkish college students: The role of psychological health, demographic, and usage characteristics. *Cyberpsychology, Behavior, and Social Networking* (2013).
[57] Silvia Kratzer and Ulrich Hegerl. 2008. Is "Internet Addiction" a disorder of its own?–a study on subjects with excessive internet use. *Psychiatrische Praxis* (2008).
[58] Kostadin Kushlev, Jason Proulx, and Elizabeth W Dunn. 2016. Silence your phones: Smartphone notifications increase inattention and hyperactivity symptoms. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems.*
[59] Daria Kuss and Mark Griffiths. 2014. *Internet Addiction in Psychotherapy.*
[60] Daria J Kuss and Mark D Griffiths. 2011. Online social networking and addiction - a review of the psychological literature. *International Journal of Environmental Research and Public Health* (2011).
[61] Daria J Kuss and Mark D Griffiths. 2017. Social networking sites and addiction: Ten lessons learned. *International Journal of Environmental Research and Public Health* (2017).
[62] Simone Lanette, Phoebe K Chua, Gillian Hayes, and Melissa Mazmanian. 2018. How Much is' Too Much'?: The Role of a Smartphone Addiction Narrative in Individuals' Experience of Use. *Proceedings of the ACM Conference on Computer Supported Cooperative Work and Social Computing* (2018).
[63] Jong-Eun Roselyn Lee, David Clark Moore, Eun-A Park, and Sung Gwan Park. 2012. Who wants to be "friend-rich"? Social compensatory friending on Facebook and the moderating role of public self-consciousness. *Computers in Human Behavior* (2012).
[64] Zach WY Lee, Christy MK Cheung, and Dimple R Thadani. 2012. An investigation into the problematic use of Facebook. In *Proceedings of the Hawaii International Conference on System Sciences.*
[65] Maxwell McCombs. 2002. The agenda-setting role of the mass media in the shaping of public opinion. In *Mass Media Economics 2002 Conference, London School of Economics.*
[66] Soraya Mehdizadeh. 2010. Self-presentation 2.0: Narcissism and self-esteem on Facebook. *Cyberpsychology, Behavior, and Social Networking* (2010).
[67] Megan A Moreno, Lauren A Jelenchick, Katie G Egan, Elizabeth Cox, Henry Young, Kerry E Gannon, and Tara Becker. 2011. Feeling bad on Facebook: Depression disclosures by college students on a social networking site. *Depression and Anxiety* (2011).
[68] Jonathan A Obar, Paul Zube, and Clifford Lampe. 2012. Advocacy 2.0: An analysis of how advocacy groups in the United States perceive and use social media as tools for facilitating civic engagement and collective action. *Journal of Information Policy* (2012).
[69] Ursula Oberst, Elisa Wegmann, Benjamin Stodt, Matthias Brand, and Andrés Chamarro. 2017. Negative consequences from heavy social networking in adolescents: The mediating role of fear of missing out. *Journal of Adolescence* (2017).
[70] Brid O'Conaill and David Frohlich. 1995. Timespace in the workplace: Dealing with interruptions. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems: Companion.*
[71] World Health Organization. 2018. International Classification of Diseases (ICD-11). https://icd.who.int/browse11/l-m/en.
[72] Wei Peng, Junghyun Kim, and Robert LaRose. 2010. Social networking: Addictive, compulsive, problematic, or just another media habit? In *A networked self.*
[73] KU Petersen, N Weymann, Y Schelb, R Thiel, and R Thomasius. 2009. Pathological Internet use–epidemiology, diagnostics, co-occurring disorders and treatment.
[74] Ronald Pies. 2009. Should DSM-V designate "Internet addiction" a mental disorder? *Psychiatry* (2009).
[75] Philip M Podsakoff, Scott B MacKenzie, and Nathan P Podsakoff. 2012. Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology* (2012).

[76] Andrew K Przybylski, Kou Murayama, Cody R DeHaan, and Valerie Gladwell. 2013. Motivational, emotional, and behavioral correlates of fear of missing out. *Computers in Human Behavior* (2013).

[77] Andrew K Przybylski and Netta Weinstein. 2017. A large-scale test of the goldilocks hypothesis: quantifying the relations between digital-screen use and the mental well-being of adolescents. *Psychological Science* (2017).

[78] Ameet Ranadive and David Ginsberg. 2018. New Tools to Manage Your Time on Facebook and Instagram. https://newsroom.fb.com/news/2018/08/manage-your-time/.

[79] James Roberts, Luc Yaya, and Chris Manolis. 2014. The invisible addiction: Cell-phone activities and addiction among male and female college students. *Journal of Behavioral Addictions* (2014).

[80] Hans-Jürgen Rumpf, Ad A Vermulst, Anja Bischof, Nadin Kastirke, Diana Gürtler, Gallus Bischof, Gert-Jan Meerkerk, Ulrich John, and Christian Meyer. 2014. Occurence of internet addiction in a general population sample: a latent class analysis. *European Addiction Research* (2014).

[81] Tracii Ryan, Andrea Chester, John Reece, and Sophia Xenos. 2014. The uses and abuses of Facebook: A review of Facebook addiction. *Journal of Behavioral Addictions* (2014).

[82] Tracii Ryan and Sophia Xenos. 2011. Who uses Facebook? An investigation into the relationship between the Big Five, shyness, narcissism, loneliness, and Facebook usage. *Computers in Human Behavior* (2011).

[83] Ashley M Shaw, Kiara R Timpano, Tanya B Tran, and Jutta Joormann. 2015. Correlates of Facebook usage patterns: The relationship between passive Facebook use, social anxiety symptoms, and brooding. *Computers in Human Behavior* (2015).

[84] We Are Social. 2018. Global Digital Report 2018. https://digitalreport.wearesocial.com/.

[85] Cheri Speier, Iris Vessey, and Joseph S Valacich. 2003. The effects of interruptions, task complexity, and information presentation on computer-supported decision-making performance. *Decision Sciences* (2003).

[86] Laurence Steinberg. 2008. A social neuroscience perspective on adolescent risk-taking. *Developmental Review* (2008).

[87] Laurence Steinberg and Kathryn C Monahan. 2007. Age differences in resistance to peer influence. *Developmental Psychology* (2007).

[88] Jih-Hsin Tang, Ming-Chun Chen, Cheng-Ying Yang, Tsai-Yuan Chung, and Yao-An Lee. 2016. Personality traits, interpersonal relationships, online social support, and Facebook addiction. *Telematics and Informatics* (2016).

[89] Ran Tao, Xiuqin Huang, Jinan Wang, Huimin Zhang, Ying Zhang, and Mengchen Li. 2010. Proposed diagnostic criteria for internet addiction. *Addiction* (2010).

[90] Eveline Teppers, Koen Luyckx, Theo A Klimstra, and Luc Goossens. 2014. Loneliness and Facebook motives in adolescence: A longitudinal inquiry into directionality of effect. *Journal of Adolescence* (2014).

[91] Jean M Twenge, Gabrielle N Martin, and W Keith Campbell. 2018. Decreases in psychological well-being among American adolescents after 2012 and links to screen time during the rise of smartphone technology. *Emotion* (2018).

[92] Recep Uysal, Seydi Ahmet Satici, and Ahmet Akin. 2013. Mediating effect of Facebook® addiction on the relationship between subjective vitality and subjective happiness. *Psychological Reports* (2013).

[93] Patti M Valkenburg and Jochen Peter. 2007. Online communication and adolescent well-being: Testing the stimulation versus the displacement hypothesis. *Journal of Computer-Mediated Communication* (2007).

[94] Patti M Valkenburg and Jochen Peter. 2009. The effects of instant messaging on the quality of adolescents' existing friendships: A longitudinal study. *Journal of Communication* (2009).

[95] Regina JJM Van den Eijnden, Gert-Jan Meerkerk, Ad A Vermulst, Renske Spijkerman, and Rutger CME Engels. 2008. Online communication, compulsive Internet use, and psychosocial well-being among adolescents: a longitudinal study. *Developmental Psychology* (2008).

[96] Philippe Verduyn, David Seungjae Lee, Jiyoung Park, Holly Shablack, Ariana Orvell, Joseph Bayer, Oscar Ybarra, John Jonides, and Ethan Kross. 2015. Passive Facebook usage undermines affective well-being: Experimental and longitudinal evidence. *Journal of Experimental Psychology: General* (2015).

[97] Erin A Vogel, Jason P Rose, Bradley M Okdie, Katheryn Eckles, and Brittany Franz. 2015. Who compares and despairs? The effect of social comparison orientation on social media use and its outcomes. *Personality and Individual Differences* (2015).

[98] Aviv Weinstein and Michel Lejoyeux. 2010. Internet addiction or excessive internet use. *American Journal of Drug and Alcohol Abuse* (2010).

[99] Barry Wellman and Scot Wortley. 1990. Different strokes from different folks: Community ties and social support. *Amer. J. Sociology* (1990).

[100] Cheng-Fang Yen, Chih-Hung Ko, Ju-Yu Yen, Yu-Ping Chang, and Chung-Ping Cheng. 2009. Multi-dimensional discriminative factors for Internet addiction among adolescents regarding gender and age. *Psychiatry and Clinical Neurosciences* (2009).

[101] Kimberly S Young. 1996. Psychology of computer use: XL. Addictive use of the Internet: a case that breaks the stereotype. *Psychological Reports* (1996).

[102] Kimberly S Young. 1998. Internet addiction: The emergence of a new clinical disorder. *Cyberpsychology & Behavior* (1998).

[103] Kimberly S Young and Cristiano Nabuco De Abreu. 2010. *Internet Addiction: A Handbook and Guide to Evaluation and Treatment.*